Pre-PhD Synopsis Report

on

Clinical Parser Optimization using Probabilistic Context Free Grammar

Under the Guidance of:

Dr. A. Senthil

Submitted by:

Anjali Kedawat

Enroll. no.:160487

Department of Computer Science and Engineering

School of Engineering and Technology

Mody University of Science and Technology, Laxmangarh

September,2020

Contents

1	Introduction	1-19
	1.1 Treebank	1
	1.2 Part of Speech Tags	3
	1.2.1 Genia Tagger	3
	1.2.2 Penn Treebank	5
	1.2.3 Type of POS	7
	1.3 Phrase Structure Grammar	8
	1.3.1 Phrases	9
	1.3.2 Clause	13
	1.4 Context Free Grammar	13
	1.4.1 CFG for Clinical Narratives	14
	1.5 Probabilistic Context Free Grammar	17
	1.6 Evaluation Criteria	18
	1.7 Research Motivation	18
	1.8 Outline	19
2	Literature Review	20-24
3	Simple Clinical Parser for clinical Narratives	25-29
	3.1 Clinical Parser	25
	3.2 Clinical Parser with the use of CFG	26
4	Parsing Clinical Narratives using Rule based Techniques	30-39
	4.1Rule based Technique in Clinical narratives	30
5	Parsing with the use of Phrase Structure Grammar	40-48
	5.1 Parsing for Clinical Narratives	40
	5.2 Phrase Structure Grammar for clinical sentence	41
6	Design of clinical parser with the use of PCFG	49-54
	6.1 PCFG for clinical Narratives	49
	6.2 Result and Discussion	50
	6.3 PCFG for neck Report	50
7	Analysis of other Reports	55-78
	7.1 USG of neck, cheek and lower abdomen	55
	7.2 USG Reports	63
	Conclusion and Future Scope	79
	References	80
	Bibliography	83
8	List of Publications	84

Chapter1

INTRODUCTION

Parsing is performed by a module in compiler which is known as parser generator. It is required for each software system. Whenever a programming language evolved it needs parser generator for parsing. Parser Generator will automatically generate a program to parse a text file into a parse tree, given a description of the language. Such software is called a "parser generator", or less accurately, a "compiler compile "It is a heart of any compiler.It takes input in form of tokens and generate output in form of parse tree or an abstract syntax tree i.e. it gives output in meaningful phrases which give output in meaningful sentences in semantic phase .Abstract syntax tree is a tree structure for representing the structure of program code. Parsing is of two types: Constituency Parsing and Dependency Parsing. Parser is also required for big data analysis which is very popular in today's scenario. Parser generator is multidisciplinary as it requires in every field which have software for performing efficient work.

1.1Treebank

Parsing is one of the components of Natural Language processing system which is used for medical domain. Natural Language Processing is used for processing of natural language like text. Phrase structure grammar is used for processing of NLP systems. There is various treebank which is used for various annotation. Table 1.1 shows various treebanks used for annotation.

Treebank	Source	Annotation scheme
Penn treebank	The material annotated includes such wide-ranging genres as IBM computer manuals, nursing notes, Wall Street Journal articles. produced approximately 7 million words of part-of-speech tagged text, 3 million words of skeletally parsed text, over 2 million words of text parsed for predicate argument structure, and 1.6 million words of transcribed spoken text	POS Tagging, Syntactic bracketing.
Brown Laboratory for	Treebank-style part-of-speech (POS) tagged and	POS tagging

Linguistic Information Processing (BLLIP) WSJ Corpus	parsed version of the three-year Wall Street Journal (WSJ) approximately 30 million words	
British Component of the International Corpus of English (ICE-GB)	One million words of spoken and written British English from the 1990s	POS tagging
Diachronic Corpus of Present-Day Spoken English (DCPSE)	It contains more than 400,000 words from ICE-GB and 400,000 words from the London-Lund Corpus. It contains 87,188 parse trees.	POS tagging
Susanne Corpus	130,000 words of written English based on various genres in the Brown Corpus that have been parsed and marked up based on an "annotation scheme"	POS tagging

Table 1.1: Treebanks

Treebank corpus. Table 1.1.1 shows syntactic treebanks have been developed for a wide variety of languages:

Language	Treebank	Syntactic Formalism
Language	Treebank	Syntactic Formalism
English	Penn Treebank	Phrase structure
English	Prague English Dependency Treebank	Dependency
English	Universal Dependencies	Dependency
English	BLLIP WSJ corpus	Phrase structure
English	British Component of the International Corpus of English (ICE-GB)	Phrase structure
English	Diachronic Corpus of	Phrase structure

	Present-Day Spoken English (DCPSE)	
English	Lancaster Parsed Corpus	Phrase structure
English	Susanne Corpus	Phrase structure
English	Christine Corpus	Phrase structure
English	Lucy Corpus	Phrase structure
English	The PARC 700 Dependency Bank	Dependency
English	CHILDES Brown Eve corpus with dependency annotation	Dependency
English	SMULTRON - Parallel Treebank EN-DE-SV	Phrase structure
English (historical)	Penn Parsed Corpora of Historical English;	Phrase structure
English (historical)	York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE)	Phrase structure
Hindi	AnnCorra	Dependency

Table 1.1.1: Treebanks used for English and Hindi domain

1.2Part of Speech Tags

POS (part of speech tags) is assigned to each word in text corpus. It is also known as grammatical parsing because it also includes verbs, noun, preposition, adjective and their sub -categories. Tagging is used to assign description to tokens i.e. from which POS that token belongs to. There are many POS tags available for medical lexicons. Some of them are as follows:

1.2.1 GENIA Tagger

It is used for named entity recognition. GENIA POS tagging is shown in Table 1.2.1

PTB	GENIA	Description
CC		Coordinating conjunction.
CD		Cardinal number.
DT		Determiner.
EX		Existential *there*.
FW		Foreign word.
IN		Preposition or subordinating conjunction.
JJ		Adjective.
JJR		Adjective, comparative.
JJS		Adjective, superlative.
LS		List item marker.
MD		Modal.
NN		Noun, singular or mass.
NNS		Noun, plural.
NNP		Proper noun, singular.
NNPS		Proper noun, plural.
CC		Coordinating conjunction.
CD		Cardinal number.
DT		Determiner.
EX		Existential *there*.
FW		Foreign word.
IN		Preposition or subordinating conjunction.
JJ		Adjective.
JJR		Adjective, comparative.
JJS		Adjective, superlative.
LS		List item marker.
MD		Modal.
NN		Noun, singular or mass.
NNS		Noun, plural.
PDT		Predeterminer.
POS		Possessive ending.
PRP		Personal pronoun.
PRP$	PRPP	Personal pronoun, possessive
RB		Adverb.
RBR		Adverb, comparative.
RBS		Adverb, superlative.
RP		Particle.
SYM		Symbol.
TO		*to*.
UH	-	Interjection. This does not appear in the GENIA corpus.
VB		Verb, base form.
VBD		Verb, past tense.
VBG		Verb, present participle or gerund.
VBN		Verb, past participle.
VBP		Verb, non-3rd person singular present.
VBZ		Verb, 3rd person singular present.
WDT		*Wh*-determiner.
WP		*Wh*-pronoun.

WP$	WPP	*Wh*-pronoun, possessive.
	WRB	*Wh*-adverb.
#	-	Pound sign. This does not appear in the GENIA corpus.
$	-	Dollar sign. This does not appear in the GENIA corpus.
.	PERIOD	Period.
,	COMMA	Comma.
:	COLON	Colon.
(	LRB	Left one of any paired symbols used as brackets: (, [, {, <.
)	RRB	Right one of any paired symbols used as brackets:),],}, >.
“	LQT	Left quotation mark, single or double
”	RQT	Right quotation mark, single or double

Table 1.2.1 GENIA POS Tags

1.2.2 Penn Treebank POS

It is used for English corpora. It contains 36 pos tags and 12 other symbols.

POS Tag	Description	Example
CC	coordinating conjunction	And
CD	cardinal number	1, third
DT	determiner	The
EX	existential there	there is
FW	foreign word	Les
IN	preposition, subordinating conjunction	in, of, like
IN/that	that as subordinator	That
JJ	adjective	Green
JJR	adjective, comparative	Greener
JJS	adjective, superlative	Greenest
LS	list marker	1)
MD	modal	could, will
NN	noun, singular or mass	Table
NNS	noun plural	Tables
NP	proper noun, singular	John
NPS	proper noun, plural	Vikings
PDT	predeterminer	both the boys
POS	possessive ending	friend’s
PP	personal pronoun	I, he, it

PPZ	possessive pronoun	my, his
RB	adverb	however, usually, naturally, here, good
RBR	adverb, comparative	Better
RBS	adverb, superlative	Best
RP	particle	give up
SENT	Sentence-break punctuation	. ! ?
SYM	Symbol	/ [= *
TO	infinitive 'to'	To go
UH	interjection	Uhhuhhuhh
VB	verb be base form	Be
VBD	verb be past tense	was, were
VBG	verb be gerund/present participle	Being
VBN	verb be past participle	Been
VBP	verb be, sing. present, non-3d	am, are
VBZ	verb be, 3rd person sing. present	Is
VH	verb have, base form	Have
VHD	verb have, past tense	Had
VHG	verb have, gerund/present participle	Having
VHN	verb have, past participle	Had
VHP	verb have, sing. present, non-3d	Have
VHZ	verb have, 3rd person sing. present	Has
VV	verb, base form	Take
VVD	verb, past tense	Took
VVG	verb, gerund/present participle	Taking
VVN	verb, past participle	Taken

VVP	verb, sing. present, non-3d	Take
VVZ	verb, 3rd person sing. present	Takes
WDT	wh-determiner	Which
WP	wh-pronoun	who, what
WP$	possessive wh-pronoun	Whose
WRB	wh-adverb	where, when
#	#	#
$	$	$
“	Quotation marks	‘ “
``	Opening quotation marks	‘ “
(	Opening brackets	({
)	Closing brackets	) }
,	Comma	,
:	Punctuation	– ; : — …

Table 1.2.2 Penn Treebank POS tag sets

1.2.3Types of POS

POS tagging is of three types Rule based, Stochastic and Transformation base tagging.

1.2.3.1 Rule Base POS tagging

Rule base POS tagging are lexicon for each word in a sentence. For each word rule is assigned according to dictionary. If for a single word more than one possibility of tagging then rule is written manually. It is done by either context free grammar or regular expression format. In context free grammar each tagger is assigned to each word. In regular expression word is written in regular expression format.

1.2.3.2 Stochastic Base POS tagging

Stochastic POS tagging is based on two approach: a) probability assigned to each word this probability assigned according to occurrence of word in particular tag .b)probability assigned to each tag according to occurrence of tag for particular production rule

1.2.3.3 Transformation base POS tagging

It is useful for automatic POS tagging. It contains the procedure of rule base tagging. The only difference is that rule base tagging is done manually while transformation base tagging is done by automatic POs tagging of text.

1.3 Phrase structure Grammar

Grammar is a set of rules by which sentence can be defined .Grammar is having four tuples(V,T,S,P).Here V represent set of variables or non-terminals represent set of terminal symbol or input symbols,S is a start symbol from which processing will be started and P is a production rules. For example = ({A, B}, {a, b}, A, P) here P is following

A->aB/b

B->bA/a

According to above grammar, it is having two nonterminal "A" and "B" two terminal symbols "a "and "b" and starting symbol is "A" from which processing will be started and production rules which is used for generating parse tree. Context free Grammar is having left side of production as a nonterminal and right side as combination of terminal and nonterminal. Figure 1.3 is an example of CFG. If input string is" abaa" then parse tree for this string is following:

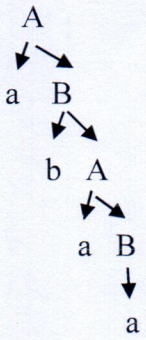


Figure 1.3: Parse tree for input" abaa"

In reference of NLP grammar is in the form of phrases so it is known as phrase structured grammar. For the sentence" Kidney has normal episodes and shape with normal dimensions" phrase structured grammar is as follows:

S->NPVP

NP->NN/NP PP/NP CC NN/JJ NNS

VP->VBZ NP

PP->IN NP

NN->Kidney/shape

VBZ->has

NNS->episodes/dimension

JJ->normal

CC->and

IN->with

In above grammar following rules are performed:

Rule 1: noun phrase should follow by verb phrase

Rule 2: Noun singular, noun phrase followed by preposition phase or noun phrase followed by Coordinating conjunction followed by Noun singular or adjective followed by noun plural

Rule 3: Verb third person followed by noun phrase

Rule 4: Preposition or subordinating conjunction followed by noun phrase

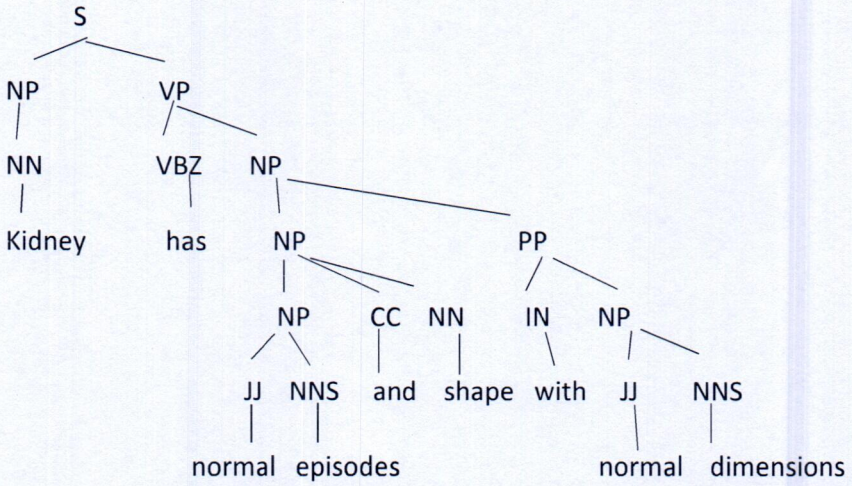


Figure 1.3.1.: Parse tree for "Kidney has normal episodes and shape with normal dimensions"

1.3.1 Phrases

Phrase is a collection of word that is used as a unit in a sentence. There are five main types of phrases: noun, verb, adverb, adjective and prepositional. In case of linguistic analysis, single word and a group of words is known as phrases that is functioning like f a constituent which helps to generate sentence syntax.

1.3.1.1 Noun Phrase

Noun can be a person, place or thing. Whenever any of that present in a sentence then noun phrase assigned to that person, place or thing. Noun can be abstract or concrete, proper or common. singular or plural, collective.

(a) Abstract or Concrete Noun

Abstract noun is a name or thing which cannot touch and see. In clinical report measurements are one of the examples of abstract noun. For example: Heart rate is 132b/s, here 132 is an abstract noun. Concrete noun is a name or things that we can touch and see. In clinical report name of patient is an example of concrete noun

(b) Proper or Common Noun

Proper noun is particular name of a person or thing. It is denoted by NNP (singular proper noun) and NNPS (plural proper noun) In clinical reports name of a particular person can be considered as proper noun. For example: Rajesh is having high blood pressure, here Rajesh is a proper noun. Common noun is not name of a particular person or thing. For example, heart rate is 132b/s, here heart is common noun.

(c) Singular or Plural Noun

Plural noun is having s or es with verb. It is denoted by NNS in tree structure for example, Kidneys. Singular noun is opposite of plural noun i.e. which is not having s or es with verb. It is denoted by NN in tree structure for example, Heart.

(d) Collective Noun

As its name implies, collection of things. For example, Liver

Noun phrase is a collection of words which is headed by a noun or in other words noun with modifiers. In case of parse tree phrase treated as an internal node and words as leaf node. Pre modifier (part of noun phrase come before noun) include determiners, quantifiers, numbers and adjectives. Determiners are denoted by DT, Quantifiers by Q, numbers by CD and adjective by JJ. Post modifier (part of noun phrase come after noun) include preposition phrase (PP), verb phrase (VP). It is denoted by NP(Noun Phrase). For example, parse tree for clinical sentence "No evidence of biliary duct dilatation" is shown in Figure 1.3.1.1

S

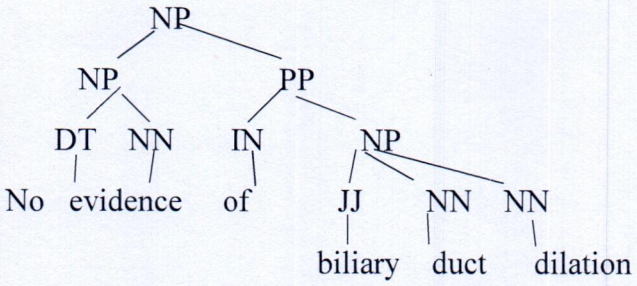


Figure 1.3.1.1 Parse tree for "No evidence of biliary duct dilatation"

1.3.1.2 Verb Phrase

Verb phrase containing verbs which works as a head word. These phrases are lexical units. It is having two types: verb itself and adjective, noun, adverb as a part of verb phrases. It is denoted by VP (verb Phrase). In Part of speech tags VBP is used for non-3rd person singular present, VBZ 3rd person singular present, VBN past participle, VBG gerund or present participle, VBD represent verb past tense. For example, parse tree for sentence" Limb movements are normal "is shown in Figure 1.3.1.2

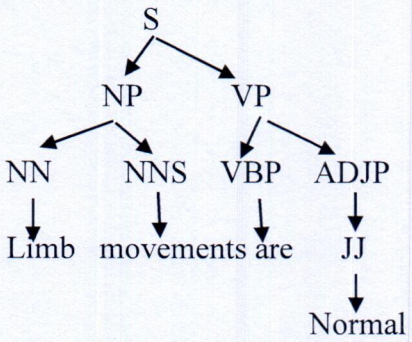
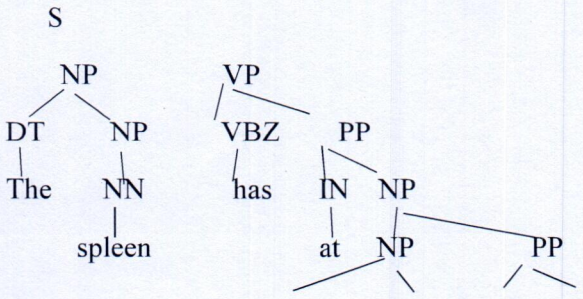


Figure 1.3.1.2 Parse tree for "Limb movements are normal"

1.3.1.3 Preposition Phrase

It is a group of words which usually starts with preposition. It mainly constitutes preposition followed by noun phrase. It is denoted by PP (preposition phrase).IN, TO represent preposition or subordinating conjunction. For example, parse tree for sentence "The spleen has at the upper limit of normal size" is shown in Figure 1.3.1.3



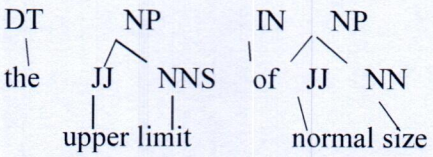


Figure 1.3.1.3 Parse tree for "The spleen has at the upper limit of normal size"

1.3.1.4 Adjective Phrase

It defines pronoun and noun in a sentence. It can come start, end or middle of phrase in a sentence. It can appear before or after of pronoun and noun in a sentence. It is denoted by ADJP (adjective phrase). JJ represent adjective, JJR adjective comparative, JJS adjective superlative. For example, parse tree for sentence "The visible vessels are normal" is shown in Figure 1.3.1.4

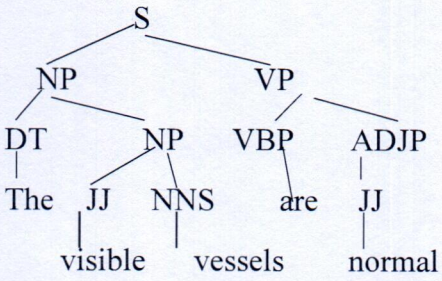


Figure 1.3.1.4 Parse tree for" The visible vessels are normal"

1.3.1.5 Adverb Phrase

It includes adverb, verb, adjectives. It is denoted by ADVP (adverb phrase). RB denotes adverb, RBR adverb comparative, RRS adverb superlative. For example, parse tree for sentence "Normal in size, shape and echotexture "is shown in Figure 1.3.1.5

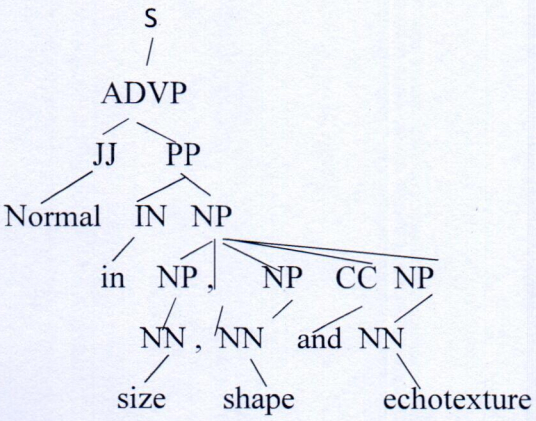


Figure 1.3.1.5 Parse tree for "Normal in size, shape and echotexture "

1.3.2 Clause

A Clause is a collection of words which contains noun phrase and verb. In other words, clause contain phrases Phrase can also be a clause and contain clause within it. There are two types of clauses:

1.3.2.1Dependent Clause

Dependent Clause starts with subordinate conjunctions like since, when, because, although etc. For example: When I went, this sentence starts with when subordinate conjunction, so it is a dependent clause.

1.3.2.2 Independent Clause

Independent clause is a collection of words that make a meaningful sentence. For example, I am eating food. Two independent clauses can be joined by semicolon, comma or subordinate conjunction. Independent clauses are strong than Dependent clauses as independent clause provide meaningful thoughts while it is not the case for dependent clause.

1.4 Context Free Grammar

Context free grammar is having 4-tuple and denoted by G= (V, T, S, P)

V is a set of nonterminal

T is set of terminal symbols

S is a starting symbol

P is a production rule

Left side of production is always single nonterminal and right side can be nonterminal or terminals or combination of both. For example, S->cS/bS/b/epsilon here S represent nonterminal, (c, b, epsilon) is terminal symbol and S is a starting symbolic .CFG can be ambiguous grammar if more than one parse tree possible for a single sentence according to given grammar. For a string "cbb" two parse tree possible from above grammar which is shown in Figure 3.1 and Figure 3.1.1

S->cS->cbS->cbb

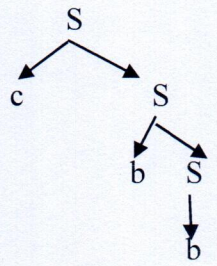


Figure 1.4. Parse tree for string "cbb"

S->cS->cbS->cbbS->cbb

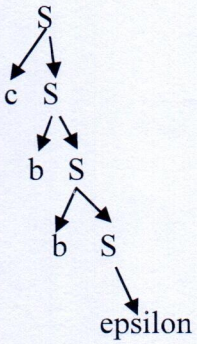


Figure 1.4.1. Parse tree for string "cbb"

So above grammar is ambiguous grammar.

1.4.1 CFG for Clinical Narratives

Clinical Narratives can be parsed by phrase structured grammar which is having similar properties of context free grammar. For clinical sentence "Both lobes of thyroid glands appear normal" following rules can be generated

S->NP VP

NP->NPPP/DT NP/NNS/NNNP

VP ->VBP ADJP

PP->IN NP

NNS->lobes/glands

NN->thyroid

DT->Both

IN->of

JJ->normal

ADJP->JJ

VBP->appears

Parse tree for above clinical sentence can be generated by two ways, so it is one of the examples of ambiguous grammar for clinical sentences.

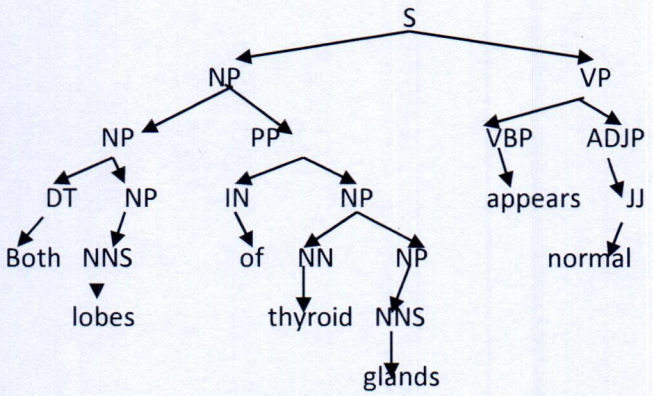


Figure 1.4.2 Parse tree for "Both lobes of thyroid glands appears normal"

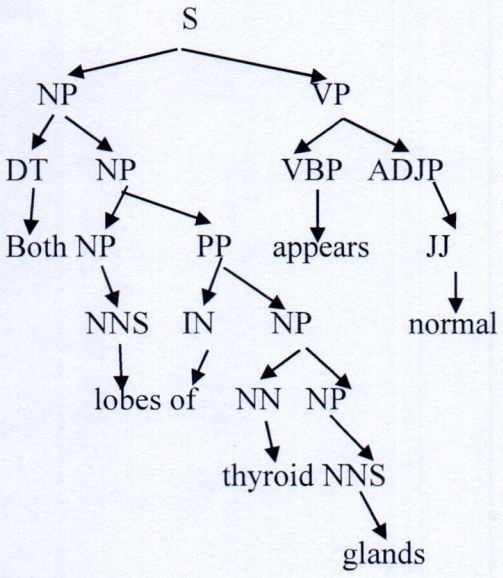


Figure 1.4.3 Parse tree for "Both lobes of thyroid glands appears normal"

Ambiguity can be removed by weighted CFG, as earlier mentioned it provides partial solution to remove ambiguity. In which weight is assigned to every production according to weight production will be selected to remove ambiguity.

To remove ambiguity from above grammar, in NP production instead of using two different production DT NP, NP PP we make it as one production like NP->DTNP/NNSPP. After this modification grammar will be

S->NP VP

NP->NNNP/DT NP/NNSPP/NNS

VP ->VBP ADJP

PP->IN NP

NNS->lobes/glands

NN->thyroid

DT->Both

IN->of

JJ->normal

ADJP->JJ

VBP->appears

Parse tree for "Both lobes of thyroid glands appears normal "is shown in Figure 1.4.4

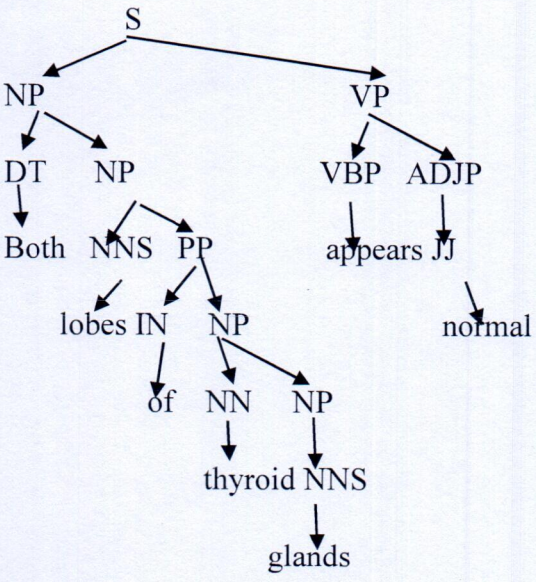


Figure 1.4.4Parse tree for without ambiguity "Both lobes of thyroid glands appears normal "

1.5 PCFG (Probabilistic Context free grammar)

PCFG (probabilistic context free grammar) is a weighted context free grammar in which probability assigned to each production. Total probability for a production rule should be 1. It eliminates ambiguity partially in other words it provides partial solution for elimination of ambiguity. PCFG is having 5-tuples (V, T, S.P.PR):

V is a set of nonterminals

T is a set of terminals

S is a starting symbol

P is a set of production rule

PR is a set of probabilities assigned to production rule.

S->NPVP 1

NP->NN .2 /NP PP .2/NP CC NN .2/JJ NNS.4

VP->VBZ NP 1

PP->IN NP 1

NN->Kidney .5/shape .5

VBZ->has 1

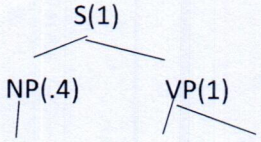
NNS->episodes .5/dimension .5

JJ->normal 1

CC->and 1

IN->with 1

In above grammar S production is having only one option so probability (1) assigned to it. In NP production rule NN and JJNNS is used more in comparison to other two production so high probability assigned to more used production and less probability to less used production



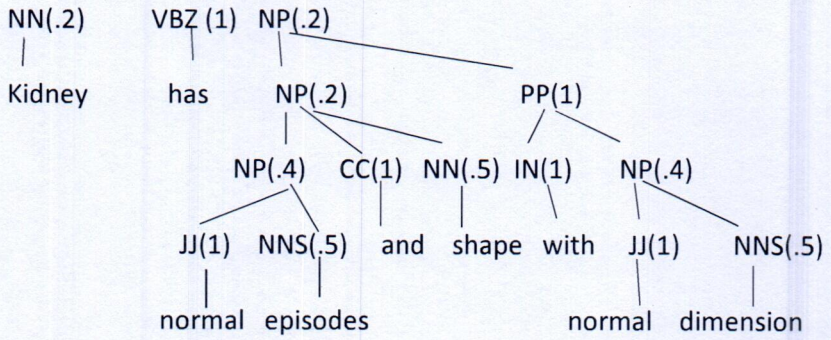


Figure 1.5 Parse tree for sentence "Kidney has normal episodes and shape with normal dimension"

Approximation of above parse tree is calculated by multiply all the probabilities like:

$1x.4x.5x1x1x.1x.1x.4x1x.5x1x.5x1x1x.4x1x.5=4x10^{-5}$

1.6 Evaluation Criteria

Accuracy of parse tree is calculated by recall, precision and bracketing f- measure

1.6.1 Bracketing Recall

Bracketing Recall (BR)= number of correctly classified constituents in system parse tree /total constitutes of gold standard parse tree

1.6.2 Bracketing Precision

Bracketing Precision (BP)= number of correctly classified constituents in system parse tree/number of all constituents in system parse tree

1.6.3 Bracketing f-measure

Bracketing f-measure=2* BR* BP/BR + BP

1.7 Research Motivation

Clinical Parsing is important for clinical reports. Many parsers generate for parsing of English domain and some parsers are used for both English and clinical domain. We generate clinical parser with no ambiguity .Noun phrase with special symbol taken as complete sentence noun phrase in other parser and ambiguity of preposition phrase also not eliminated ,so we try to do clinical parsing with taking special symbol as part of parsing .Our parser is based on Constituency parsing approach with the use of PCFG (probabilistic context free grammar). We use stochastic POS tagging based on tagger probability approach i.e.

probability is assigned to each tag for particular production rule. In this we use ultrasound reports of neck, cheeks and general ultrasound. For each report grammar rule is different due to multilingualism in clinical narratives. These reports are taken by a hospital only as different hospital is having different format of report for the same kind of disease. Approximation of each sentence is also calculated and then recall; precision and bracketing f-measure is calculated with comparison of Sandford parser which is well known parser for both lexicalized and unlexicalized parsing. But it is not accurate for clinical narratives as it assumes noun phrase after normal sentence as a big noun phrase, ambiguity in prepositional phrase, symbols of clinical sentence can not read by this parser. From our parser we can easily check it is how much accurate by seeing recall precision and bracketing f-measure. Constituency parser is preferable to dependency parser because it is easier to understand for anyone due to grammar tagging, it not only displays relations but also provide structure of sentence. Disambiguous grammar is generated so that only from one rule of grammar parsing can be done. Constituency parser is based on bottom up approach. Stochastic POS tagging is manually approach so we done manual POS tagging for these reports.

1.8 Outline

The rest chapters of this report organized as

Chapter 2 provides a literature survey on general parsers which is used for clinical and non-clinical purpose and nlp systems.

Chapter 3 define first objective of our research. It describes clinical parsing used for clinical narratives with the help of clinical sentences

Chapter 4 fulfil the second objective of pour research. It defines Cconstituency Parser and Rule based Techniques for Clinical Narrative.

Chapter 5 define third objective our research. It gives phrase structured grammar with probability to improve performance of parser.

Chapter 6 describes fourth objective of our research. It gives how to design parse tree without ambiguity and output shows parse tree for clinical sentence.

Chapter 7 describe analysis of real reports. We take various types of ultrasound reports and design parse tree for clinical narratives to make it as a meaningful clinical sentence.

Chapter 8 defines conclusion of this research and future scope for this.

Chapter 2

Literature Review

Friedman *et al.* [1] build MedLEE (Medical Language Extraction and Encoding) NLP system which is used for decision support. This system initially applied to radiology reports of the chest X-ray report. The first extension of MedLEE is mammography report due to language similarity of chest X-ray report. Second extension of MedLEE is discharge summaries. This system written in PROLOG which support Windows, UNIX and AIX platform. MedLEE have four components: pre-processor which reads original clinical text, parser which is used to determine structure of sentence, phrase regularization and encoder.

Koehler and B. Spencer [2] develop an SymText NLP system which combine syntax and semantic approach for encoding of free text. It uses Bayesian n/w and used for chest x-ray report. This system achieves 86.1% recall and 85.6% precision. It contains a parser which have augmented transition network grammar and transformed grammar.

Lee M. Christensen *et al.* [3] develop MPLUS (A Probabilistic Medical Language Understanding System) NLP system which is an analysis tool for medical text and based on Bayesian n/w. This system defines relations with in a medical domain. Bayesian n/w is same as semantic n/w in respect of structure. MPLUS is differ from SymText in terms of size and modularity. It is based on bottom up Chart parser and coded in LISP with some C routines for BN's access.

Brian Hazlehurst *et al.* [4] describe MediClass NLP system which process free text and coded data for automatic detection of clinical events in EMR (Electronic Medical Records). This system is used for clinical operations and healthcare services. It is an open source system which use informatic techniques. It is also used for research projects based on NLP techniques for automatic indexing and coding of clinical contents. It uses both NLP techniques and Knowledge based techniques. This system applied on patient safety, disease surveillance and care quality.

S. Goryachev *et al.* [5] develop HITEx (Health Information Text Extraction) open source NLP system and based on data driven approach. It is used for medical and non-medical domain. It builds from General Architecture for Text Engineering framework (GATE). It accesses clinical information like diagnosis, discharge medication, smoking status, etc.

Guergana K Savova *et al.* [6] build cTAKES (clinical Text Analysis and Knowledge Extraction System) open source NLP system for extract information from EMR. It is built from open source technique (UIMA) and open NLP process toolkit. The goal of developing this system is to develop heterogeneous clinical domain system. cTAKES include sentence boundary, detector, tokenizer, normalizer, POS, shallow parser, named entity recognition. It integrates both rule based and machine learning techniques to annote negation context, named entity and syntactic context. It accepts both plain text and clinical documents.

Alan R. Aronson and François-Michel Lang [7] gives MetaMap NLP system which is used for biomedical text retrieval. The limitation of this system is that it only applies on English text and not use for real time. It reduces accuracy in presence of ambiguity. It also used for text mining, classification, Knowledge discovery etc.

Hua Xu *et al.* [8] build MedEx (Medical Information Extraction System) NLP system which is used for discharge summary. There searcher collects 50 discharge summaries with f- measure of 93.2% and 95% over 25 visit notes. It is based on semantic approach. The researcher achieves f-measure of 90% by integrating semantic tagger and chart parser. It is based on algorithm which is used to find medical name and signature. The researcher used new method of parsing by integrate sequential tagger and combined tagger. Sequential tagger is used for combination of regular expression, look up and rule-based disambiguation.

Vijay Garla *et al.* [9] develop YTEX (Yale cTAKES) NLP system which is an extension of cTAKES NLP system. It uses regular expression based named entity recognition. The researcher randomly selects 236 patient forms. It simplifies feature extraction and develop machine learning, rule-based classifiers.

Yu-Kai Lin *et al.* [10] build MedTime NLP system which is based on rule based and machine learning approach. It is a temporal information extraction system which is an important area for biomedical and health information research.

Sunghwan Sohn *et al.* [11] develop MedXN (Medical Extraction and Normalization) NLP system which is used to extract and normalize medical information by using RxNorm. It is based on UIMA framework. The researcher collects 397 medication and achieve f-score of 97.8%. The researcher used 59 clinical notes from Mayo Clinic and contained 659 manually annotated clinical mentions. It uses decomposition and composition strategy for extraction of clinical information after this MEDXN convert this information in RxCUI.

Foster R. Goss *et al.* [12] build MTERMS (Medical Text Extraction Reasoning Mapping System) NLP system to extract and encode allergy information. It converts free text into XML output by using pipeline approach. It has f-measure of 87.6%, recall 91% and precision 84.45%. The lexicon used in this system have multiple standard technologies (RxNorm, UNII, SNOMED CT).

Hua Xu et al. [13] select 50 sentences from clinical corpus and apply full parsing. Stanford parser gives F-score of 77% on the gold standard while full parsing achieves maximum BF of 81%. There are many medical systems which is used for combining semantic and syntactic rules for parsing of clinical text some of them are MedLEE (Medical Language Extraction and Encoding System), Medical Language Processing (MLP) system from the Linguistic String Project, and SymText/MPlus. Bikel parser, Collins parser, Charniak parser are used for lexicalized parsing. In lexicalized parsing parent node of each parse tree is annoted by lexeme of its main child node. Unlexicalized parsing is first introduced by Stanford parser. its performance has improved considerably by A*search Technique. Stanford parser is very useful for different domains because of its unlexicalized parsing technique. It is also used for biomedical literature.

Finkel et al. [14] describe machine learning system for named entity recognition. Stanford Parser generated syntactic information is applied on biomedical named entity recognition. They describe machine learning system for named entity recognition for five named entity of GENIA corpus. Parsing is done by unlexicalized Stanford parser.

Fundel et al. [15] describe relation extraction system which is used for biomedical text. Dependency parse trees are used for recognize these biomedical texts which is based on Stanford parser. Dependency between words which is far apart in sentences is defined by dependency parse trees which provide parse non-local dependencies. This system focus on genetic, physical and regulatory relations between genes and proteins. They give hypothesis that entities which define together are related to each other somehow. Lexicalized Stanford parser and POS tags are used for generating dependency parse trees for each sentences Medical Domain Researchers use Stanford Parser for processing of clinical text.

Huang et al. [16] describe performance of Stanford Parser for noun phrase identification in radiology reports. They apply Natural Language processing on set of radiology reports for extraction of noun phrases with full phrase. The efficiency of noun phrases in radiology

reports is judged by UMLS Specialist Lexicon. It will not include mapping of noun phrases in parse tree into corresponding UMLS concepts.

Bikel [17] used an Expectation-Maximization Model for estimation of feature space parameters of Collins model. The Bikel parser achieve high F-measure in compare to the Collins for feature space parameters that it tested. The Bikel parser was a powerful and useful imitation of the Collins parser.

Charniak and Johnson [18] construct high performance statistical parser by using discriminative re-ranking method. They select the best parse tree from the sets of 50-best parse trees. This best parse tree is used as an input to Maximum Entropy re-ranker. They achieve 91.0% F-measure which is best performance in all previous generative models.

McClosky et al. [19] presented parser which have more features than Charniak Parser This parser has two-phase parser which use Charniak parser as a first phase and a bootstrapping method as a second phase. Charniak Parser achieve F -measure of 91% and McClosky parser improves F-measure of Charniak parser by 0.8% (F-measure)

Rohit J Kate [20] presented unsupervised grammar (not restricted to binary form) in which grammar is automatically included for the sublanguage of clinical reports with no annotations. method use gold standard which is not useful for end applications and require lot of manual efforts. For example, the conventual gold method not give similarity between syntactic parse trees and semantic tree which degrade the performance of semantic parsing. Unsupervised parsing is best suited for. It creates its own non-terminals based on the application while other methods force user to specify how many non-terminals they need.

Gabriel J. Ferrer [21] is an interactive parser generator which uses Earley Algorithm. The user can edit and test grammar in an GUI interface.

Alessandro Warth et al. [22] doesn't allow semantic action written directly in grammar. This grammar is platform independent. If there is any change in previous input then Ohm parser parsed entire input that is time consuming and error-prone too.

Ricky K. Taira [23] develop a natural language processor for radiology reports. They define deficiencies of symbolic methods over statistical natural language processor. The goal of parser module is to create a dependency diagram between words in input sentences. They

describe methodology in three ways: first collect a large number of documents from domain of interest, second create the training data by manually indicating the dependency graph, third for each word in sentence define resonance condition occurred or not. It achieves recall of 90% and precision of 89%.

Friedrich Steinmann [24] present a Prolog based definition and implementation of dependency grammar that can parser sentences of a medical languages.

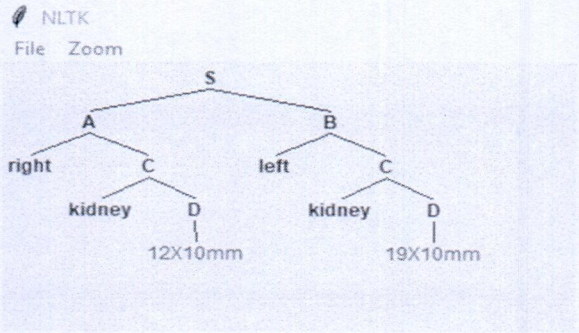
Ernestina Menasalvas et al. [25] analyze several tools and frameworks to extract medical entities by applying NLP with NER process,

P.O. El Guedj et al. [26] build Chart parser which is used to analyze large medical corpora.

Nuala A. Bennett et al. [27] extract noun phrase from Medline using general parser.

Ming Jiang, Jana Diesner [28] Define simple rule-based technique from abstracts of scientific abstracts. They provide method which automatically extracts and constructs graphs in which nodes represent text-based entities (in this case, noun phrases) and their relationships (in this case, verb phrases or preposition phrases).

```
(S
  (A (right ) (C (kidney ) (D 12X10mm)))
  (B (left ) (C (kidney ) (D 19X10mm))))
input is
['12X10mm', '19X10mm']
5
```



3.2.2 Example 2

From heart rate checking equipment gives only heart rate as a input.

Input:132

Output Heart rate 132 beats/minute

S->AB

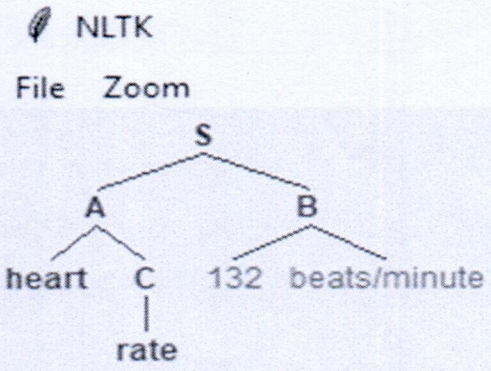
A->heart C

C->rate

B->132 beats/minute

3.2.2.1 Result and Discussion

```
(S (A (heart ) (C (rate ))) (B 132 beats/minute))
input is
['132', 'beats/minute']
4
```



3.2.3 Example 3

Input: lesion 14mm

Output: cystic lesion 14mm lobe of liver

S->AB

A->cystic C

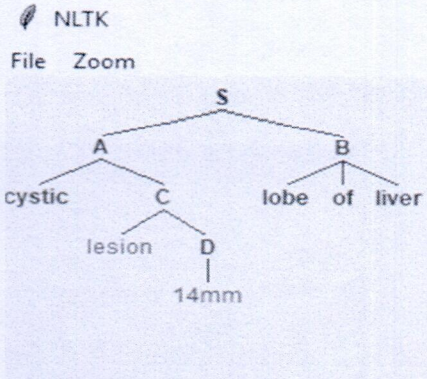
C->lesion D

D->14mm

B->lobe of liver

3.2.3.1 Result and Discussion

```
(S (A (cystic ) (C lesion (D 14mm))) (B (lobe ) (of ) (liver )))
input is
['lesion', '14mm']
5
```



3.2.4 Example 4

From ultrasound report of ovary, size is enlarged so medical equipment only gives size

Input: 18mm

Output: Enlarged size 18mm

S->A

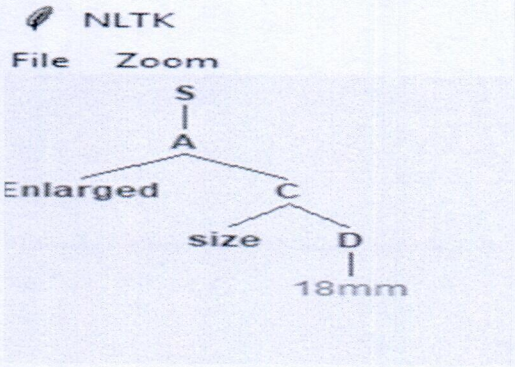
A->Enlarged C

C->size D

D->18mm

3.2.4.1 Result and Discussion

```
(S (A (Enlarged ) (C (size ) (D 18mm))))
input is
['18mm']
5
```



When it compares with gold standard then we can generate recall, precision and bracketing f-measure. If number of sentences is more than no of production also more in case of simple clinical Parser, so we improve it with the use of rule-based techniques.

Chapter 4

Objective 2: To improve the performance of parsing-clinical narratives by using rule-based techniques

Grammar is a set of rules which is used to parse a sentence. It contains five-tuples (V, E, S, P) here V represent set of nonterminal which is on left side of production and in capital or italic letters, E represent set of terminal symbols, S is starting symbol from which parsing will be started and P is production rules. Phrase structured grammar constitutes phrases as an internal node and words as a leaf node. Clinical parsing can be done by constituency parser and dependency parser. Constituency parser use phrase structure grammar for parsing of clinical sentence with the support of NLP. Constituency parse tree is having phrases as internal node and words as a leaf node each sentence. It provides visual presentation of each sentence. This parser can be created by many algorithms but here we create it by phrase structured grammar. Syntactic structure of a sentence can be defined by both parsers.

4.1 Rule base Technique in clinical narratives

Rule based parser uses handwritten rules. For creating meaningful clinical sentence, we have to create rules that is known as rule-based techniques. With this technique phrases are attached with measurements to make it as complete clinical sentence.

4.1.1 Example 1

Input: Right kidney 12x10mm Left Kidney 19x20mm

Output: Size of right kidney 12x10mm and left kidney 19x20mm

Output generated by objective1 can be refined by using rules which create a structured sentence. First, we require Size which is noun singular so it belongs to noun phrase. First rule we write for noun phrase

Rule 1: singular noun followed by noun phrase or preposition phrase followed by noun phrase or adjective followed by noun phrase or conjunction followed by noun phrase or cardinal number followed by noun phrase or singular noun

After 'Size' token 'of' token is coming which belongs to preposition phrase so we write second rule for preposition phrase

Rule 2: preposition followed by noun phrase

Stanford parser rules for same sentence is as follows:

Rule 1: noun phrase followed by conjunction followed by verb phrase or noun phrase followed by preposition phrase or singular noun or adjective followed by singular noun followed by cardinal number followed singular noun

Rule 2: preposition followed by noun phrase

Rule 3: verb followed by noun phrase

4.1.1.1 Result and Discussion

Size of right kidney 12x10 mm and left kidney 19x20 mm

Token 'Size' is singular noun which is followed by token 'of'. There are three rules of noun phrase which starts from singular noun (singular noun followed by noun phrase or singular noun). So, we cannot use only singular noun rule. So, skip that possibility, we have only one possibility left singular noun followed by noun phrase. For the same Stanford parser not giving optimized solution as noun phrase is replaced by 5 nonterminals while we replace it only 2 nonterminals for that purpose we use recursion. Stanford parser takes 'left' as verb while it is adjective. Parsing according to our rules is as follows:

Noun Phrase

Singular noun Noun phrase

Size **preposition phrase** noun phrase

Size **preposition noun phrase** noun phrase

Size of **noun phrase** noun phrase

Size of **adjective** noun phrase noun phrase

Size of right **noun phrase** noun phrase

Size of right **singular noun** noun phrase noun phrase

Size of right kidney 12x10 **singular noun** noun phrase

Size of right kidney 12x10 mm **noun phrase**

Size of right kidney 12x10 mm **conjunction** noun phrase

Size of right kidney 12x10 mm and **noun phrase**

Size of right kidney 12x10 mm and **adjective** *noun phrase*

Size of right kidney 12x10 mm and left noun ***phrase***

Size of right kidney 12x10 mm and left **singular noun** noun phrase

Size of right kidney 12x10 mm and left kidney **noun phrase**

Size of right kidney 12x10 mm and left kidney **cardinal number** noun phrase

Size of right kidney 12x10 mm and left kidney 19x20 ***noun phrase***

Size of right kidney 12x10 mm and left kidney 19x20 **singular noun**

Size of right kidney 12x10 mm and left kidney 19x20 mm

4.1.1.2 Stanford Parser Parsing

Noun Phrase

Noun Phrase Conjunction Verb Phrase

Noun Phrase Preposition Phrase Conjunction Verb Phrase

Singular Noun Preposition Phrase Conjunction Verb Phrase

Size **preposition phrase** Conjunction Verb Phrase

Size **preposition** noun phrase Conjunction Verb Phrase

Size of **noun phrase** Conjunction Verb Phrase

Size of **adjective** singular noun cardinal number singular noun Conjunction Verb Phrase

Size of right **singular noun** cardinal number singular noun Conjunction Verb Phrase

Size of right kidney **cardinal number** singular noun Conjunction Verb Phrase

Size of right kidney 12x10 **singular noun** Conjunction Verb Phrase

Size of right kidney 12x10 mm **Conjunction** Verb Phrase

Size of right kidney 12x10 mm and **Verb Phrase**

Size of right kidney 12x10 mm and **verb** noun phrase

Size of right kidney 12x10 mm and left **noun phrase**

Size of right kidney 12x10 mm and left **singular noun** cardinal number singular noun

Size of right kidney 12x10 mm and left kidney **cardinal number** singular noun

Size of right kidney 12x10 mm and left kidney 19x10 **singular noun**

Size of right kidney 12x10 mm and left kidney 19x10 mm

4.1.2 Example 2

Input: Heart rate 132 beats/minute

Output: Heart rate is 132 beats/minute

Rule 1: singular noun followed by noun phrase or cardinal number followed by noun phrase or plural noun followed by symbol followed by singular noun or singular noun

After noun verb 'is' coming so verb phrase will be introduced

Rule 2: present verb followed by noun phrase

'132' is cardinal number which is already defined in noun phrase (cardinal number followed by noun phrase) which is followed by beats that is plural noun which can be replacement of noun phrase.

Stanford Parser rules:

Rule 1: singular noun followed by singular noun or noun phrase followed by preposition phrase or cardinal number followed by plural noun or singular noun

Rule 2: verb followed by noun phrase

Rule 3: symbol followed by noun phrase

4.1.2.1 **Result** and **Discussion**

Noun phrase verb phrase

singular noun noun phrase verb phrase

Heart **noun phrase** verb phrase

Heart **singular noun** verb phrase

Heart rate **present verb** noun phrase

Heart rate is **noun phrase**

Heart rate is **cardinal number** noun phrase

Heart rate is 132 **plural noun** sym singular noun

Heart rate is 132 beats **sym** singular noun

Heart rate is 132 beats/ **singular noun**

Heart rate is 132 beats/minute

4.1.2.2 Stanford Parser parsing

Stanford parser takes symbol in preposition phrase which is error prone. While in preposition phrase, preposition is followed by noun phrase

Noun phrase verb phrase

singular noun singular noun verb phrase

Heart **singular noun** verb phrase

Heart rate **verb phrase**

Hear rate **verb** noun phrase

Heart rate is **noun phrase** preposition phrase

Heart rate is **cardinal number** plural noun preposition phrase

Heart rate is 132 **plural noun** preposition phrase

Heart rate is 132 beats **preposition phrase**

Heart rate is 132 beats **symbol** noun phrase

Heart rate is 132 beats / **noun phrase**

Heart rate is 132 beats / **singular noun**

Heart rate is 132 beats/minute

4.1.3 Example 3

Input: cystic lesion 14mm lobe liver

Output: There is evidence of cystic lesion of size 14mm in left lobe of liver

'There' belongs to EX tag so it belongs to noun phrase

Rule 1: Existential or singular noun followed by noun phrase or singular noun followed by preposition phrase or adjective followed by noun phrase or cardinal number followed by noun phrase or preposition phrase followed by noun phrase or singular noun

It is followed by verb'is' so verb phrase is as follows:

Rule 2: present verb followed by noun phrase

It is followed by 'evidence' which is singular noun.so noun phrase will be replaced by singular noun followed by noun phrase.

Rule 3: preposition followed by noun phrase

Stanford Parser Rules:

Rule 1: existential or noun phrase followed by preposition phrase followed by preposition phrase or noun phrase followed by preposition phrase or adjective followed by singular noun or nominal followed by singular noun or singular noun followed by cardinal number or singular noun

Rule 2: verb followed by noun phrase

Rule 3: preposition followed by noun phrase

4.1.3.1 Result and Discussion

Noun phrase verb phrase

Existential verb phrase

There **present verb** noun phrase

There is **noun phrase**

There is **singular noun** noun phrase

There is evidence **noun phrase**

There is evidence **preposition phrase** noun phrase

There is evidence **preposition** noun phrase noun phrase

There is evidence of **noun phrase** noun phrase

There is evidence of **adjective** noun phrase noun phrase

There is evidence of cystic **noun phrase** noun phrase

There is evidence of cystic **singular noun** noun phrase noun phrase

There is evidence of cystic lesion **noun phrase** noun phrase

There is evidence of cystic lesion **preposition phrase** noun phrase noun phrase

There is evidence of cystic lesion preposition noun phrase noun phrase noun phrase

There is evidence of cystic lesion of **noun phrase** noun phrase noun phrase

There is evidence of cystic lesion of singular **noun** noun phrase noun phrase

There is evidence of cystic lesion of size **noun phrase** noun phrase

There is evidence of cystic lesion of size **cardinal number** noun phrase noun phrase

There is evidence of cystic lesion *of size* 14 **noun phrase** noun phrase

There is evidence of cystic lesion *of size 14* **singular noun** noun phrase

There is evidence of cystic lesion *of size 14 mm* **noun phrase**

There is evidence of cystic lesion *of size 14 mm* **preposition phrase** noun phrase

There is evidence of cystic lesion *of size 14 mm* **preposition** noun phrase noun phrase

There is evidence of cystic lesion *of size 14 mm in* **noun phrase** noun phrase

There is evidence of cystic lesion *of size 14 mm in* **adjective** noun phrase noun phrase

There is evidence of cystic lesion *of size 14 mm in left* noun phrase noun phrase

There is evidence of cystic lesion *of size 14 mm in left* ***singular noun*** *noun phrase*

There is evidence of cystic lesion *of size 14 mm in left* **noun phrase**

There is evidence of cystic lesion *of size 14 mm in left* **singular noun** preposition phrase

There is evidence of cystic lesion *of size 14 mm in left lobe* **preposition phrase**

There is evidence of cystic lesion *of size 14 mm in left lobe* **preposition** noun phrase

There is evidence of cystic lesion *of size 14 mm in left lobe of* **noun phrase**

There is evidence of cystic lesion *of size 14 mm in left lobe of* **singular noun**

There is evidence of cystic lesion *of size 14 mm in left lobe of liver*

4.1.3.2 Stanford Parser Parsing

Stanford parser takes 'left word sometimes as adjective and sometime as verb

Noun phrase verb phrase

Existential verb phrase

There **present verb** noun phrase

There is **noun phrase**

There is **noun phrase** preposition phrase preposition phrase

There is **singular noun** preposition phrase preposition phrase

There is evidence **preposition phrase** preposition phrase

There is evidence **preposition** noun phrase preposition phrase

There is evidence of **noun phrase** preposition phrase

There is evidence of **noun phrase** preposition phrase preposition phrase

There is evidence of a**djective** singular noun preposition phrase

There is evidence of cystic **singular noun** preposition phrase preposition phrase

There is evidence of cystic lesion **preposition phrase** preposition phrase

There is evidence of cystic lesion **preposition** noun phrase preposition phrase

There is evidence of cystic lesion of **noun phrase** preposition phrase

There is evidence of cystic lesion of **nominal** singular noun preposition phrase

There is evidence of cystic lesion of **singular noun** cardinal number singular noun preposition phrase

There is evidence of cystic lesion of size **cardinal number** singular noun preposition phrase

There is evidence of cystic lesion of size 14 **singular noun** preposition phrase

There is evidence of cystic lesion of size 14 mm **preposition phrase**

There is evidence of cystic lesion of size 14 mm **preposition** noun phrase

There is evidence of cystic lesion of size 14 mm in **noun phrase** preposition phrase

There is evidence of cystic lesion of size 14 mm in **adjective** singular noun preposition phrase

There is evidence of cystic lesion of size 14 mm left **singular noun** preposition phrase

There is evidence of cystic lesion of size 14 mm left lobe **preposition phrase**

There is evidence of cystic lesion of size 14 mm left lobe **preposition** noun phrase

There is evidence of cystic lesion of size 14 mm left lobe of **noun phrase**

There is evidence of cystic lesion of size 14 mm left lobe of **singular noun**

There is evidence of cystic lesion of size 14 mm left lobe of liver

4.1.4 Example 4

Input: Enlarged size 18mm

Output: Enlarged in size 18mm

So as previous examples sentence starts with noun phrase so first rule belongs to noun phrase

First token is 'enlarged' which is adjective so noun phrase replaced by (adjective followed by noun phrase)

Rule 1: adjective followed by noun phrase or preposition phrase followed by noun phrase or singular noun or cardinal number followed by singular noun

As 'Enlarged' followed by 'in' which belongs to preposition phrase so second rule we will write for preposition phrase.

Rule 2: preposition followed by noun phrase

Stanford Parser Rules:

Rule 1: singular noun or singular noun followed by cardinal number

Rule 2: verb followed by preposition phrase followed by noun phrase

Rule 3: preposition followed by noun phrase

4.1.4.1 Result and Discussion

Noun phrase

Adjective noun phrase

Enlarged **noun phrase**

Enlarged **preposition phrase** noun phrase

Enlarged **preposition** noun phrase noun phrase

Enlarged in **noun phrase** noun phrase

Enlarged in **singular noun** noun phrase

Enlarged in size **noun phrase**

Enlarged in size **cardinal number** noun phrase

Enlarged in size 18 **noun phrase**

Enlarged in size 18 **singular noun**

Enlarged in size 18 mm

4.1.4.2 Stanford Parser Parsing

Stanford parser takes 'enlarged' as verb but according to sentence it should be adjective

Verb Phrase

Verb preposition phrase noun phrase

Enlarged **preposition phrase** noun phrase

Enlarged **preposition** noun phrase noun phrase

Enlarged in **noun phrase** noun phrase

Enlarged in **singular noun** cardinal number noun phrase

Enlarged in size **cardinal number** noun phrase

Enlarged in size 18 **singular noun**

Enlarged in size 18 mm

To improve this parsing, we generate graphical presentation in the form of parse tree with the use of phrase structured grammar. As rule-based techniques give meaningful clinical sentence but its time consuming so we improvise it by phrase structure grammar which gives graphical representation of clinical sentence.

Chapter 5

Objective 3: To Improvise the parser with the use of phrase structure grammar

Phrase structure grammar written in phrases i.e. noun phrase, verb phrase, preposition phrase, adjective phrase etc. and when syntactic structure will be created internal node represent phrases and external node represent words. But this grammar creates ambiguity i.e. from a single grammar and single sentence more than one syntactic structure possible then it is known as ambiguity. Stanford parser creates ambiguity because it is generated by phrase structure grammar

5.1 Parsing for clinical narratives

For example, we want to generate sentence "Limb movements are normal "so phrase structure grammar for creating this sentence is as follows:

S->NPVP

NP->NNNNS

VP->VBPADJP

NN->limb

NNS->movements

VBP->are

ADJP->JJ

JJ->normal

Parse tree generated by above grammar to generate sentence is shown in Figure 5.1

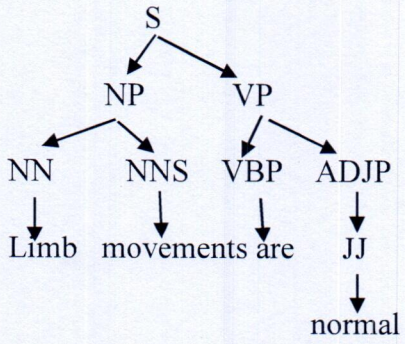


Figure 5.1 Parse Tree for sentence "Limb movements are normal"

Constituency parsing can be done by phrase structured PCFG (probabilistic context free grammar). In which probability is assigned to each production of each tagger.

5.2 Phrase structured grammar for clinical sentence

5.2.1 Example 1

Input:

Rule 1: singular noun followed by noun phrase or preposition phrase followed by noun phrase or adjective followed by noun phrase or conjunction followed by noun phrase or cardinal number followed by noun phrase or singular noun

Rule 2: preposition followed by noun phrase

Phrase structure grammar for above rules is as follows:

S->NP

NP->NNNP|PPNP|JJNP|CCNP|CDNP|NN

PP->INNP

IN->of

NN->size | mm| kidney

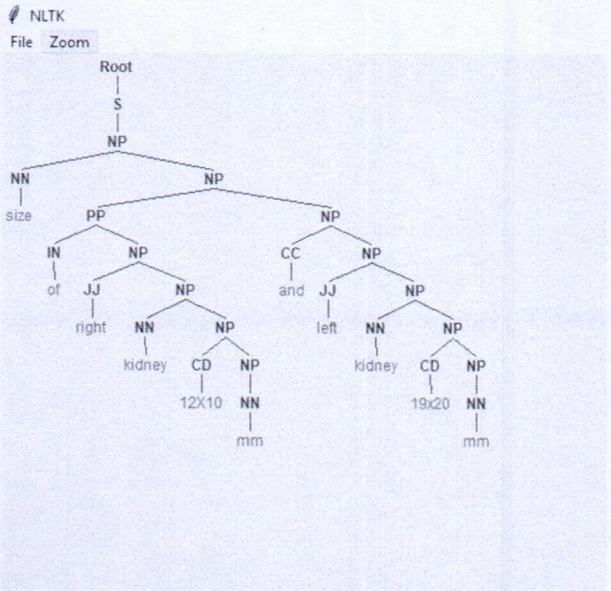
JJ->right | left

CC->and

CD->12x10|19x20

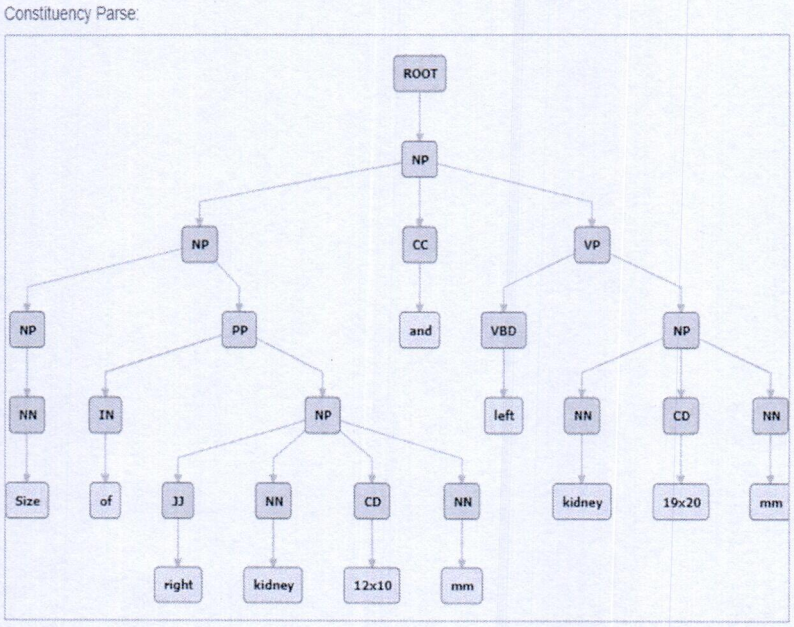
5.2.1.1 Result and Discussion

```
============= RESTART: C:\Users\hp\Desktop\phd presentation\ob31.py ============
Grammar with 15 productions (start state = S)
    S -> NP
    NP -> NNNP
    NP -> PPNP
    NP -> JJNP
    NP -> CCNP
    NP -> CDNN
    NP -> NN
    PP -> INNP
    IN -> of
    NN -> size
    NN -> mm
    NN -> kidney
    JJ -> right
    JJ -> left
    CC -> and
(Root
  (S
    (NP
      (NN size)
      (NP
        (PP
          (IN of)
          (NP
            (JJ right)
            (NP (NN kidney) (NP (CD 12X10) (NP (NN mm))))))
        (NP
          (CC and)
          (NP
            (JJ left)
            (NP (NN kidney) (NP (CD 19x20) (NP (NN mm)))))))))))
['size', 'of', 'right', 'kidney', '12X10', 'mm', 'and', 'left', 'kidney', '19x20
', 'mm']
ll
```



5.2.1.2 Comparison with Stanford parser

In it 'left' word treated as verb while it is adjective



5.2.2. Example 2

Input:

Rule 1: singular noun followed by noun phrase or cardinal number followed by noun phrase or plural noun followed by symbol followed by singular noun or singular noun

Rule 2: present verb followed by noun phrase

Phrase structure Grammar for above rules

S->NP

N->NNNP|CDNP|NNS SYM NN|NN

VP ->VBZNP

NN ->heart | rate| minute

NNS ->beats

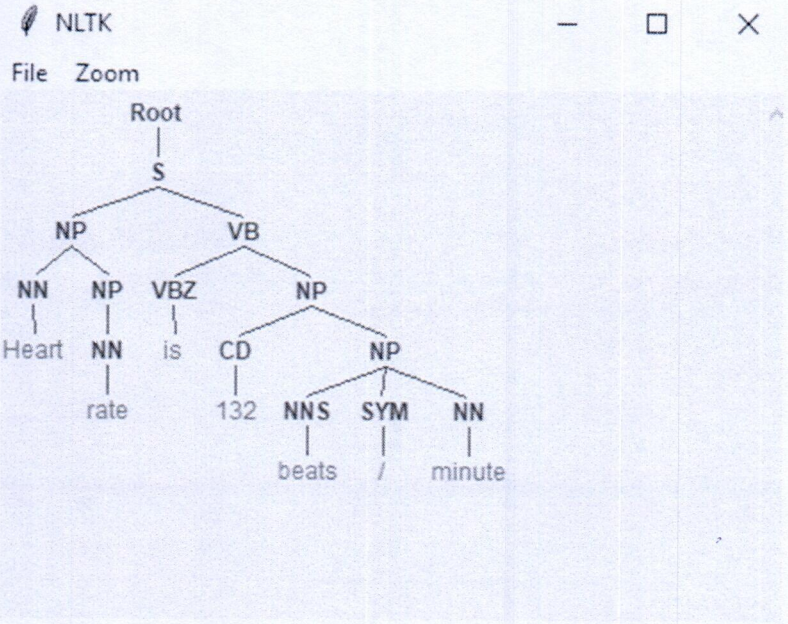
SYM -> /

VBZ ->is

CD ->132

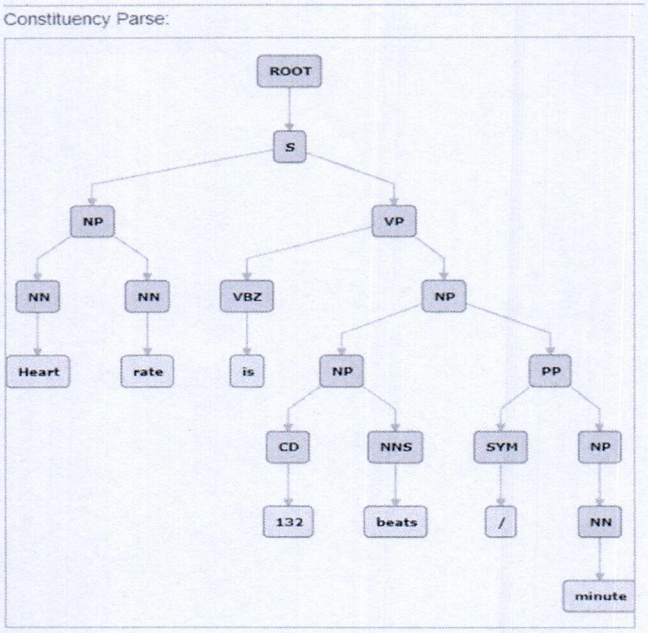
5.2.2.1 Result and Discussion

```
============= RESTART: C:\Users\hp\Desktop\phd presentation\ob32.py ======
Grammar with 13 productions (start state = S)
    S -> NP
    NP -> NN NP
    NP -> CD NP
    NP -> NNS SYM NN
    NP -> NN
    VP -> VBZ NP
    NN -> heart
    NN -> rate
    NN -> minute
    NNS -> beats
    SYM -> /
    VBZ -> is
    CD -> 132
(Root
  (S
    (NP (NN Heart) (NP (NN rate)))
    (VB (VBZ is) (NP (CD 132) (NP (NNS beats) (SYM /) (NN minute))))))
['Heart', 'rate', 'is', '132', 'beats', '/', 'minute']
7
```



5.2.2.2 Compare with Stanford Parser

Stanford parser tree height is 8 while for our parser it is 7.



5.2.3 Example 3

Input:

Rule 1: Existential or singular noun followed by noun phrase or singular noun followed by preposition phrase or adjective followed by noun phrase or cardinal number followed by noun phrase or preposition phrase followed by noun phrase or singular noun

Rule 2: present verb followed by noun phrase

Phrase structure Grammar for above rules is as follows:

S ->NPVP

NP -> EX|NNNP|NNPP|JJNP|CDNP|PPNP|NN

NN ->evidence |size |lesion |mm |left |lobe | liver

JJ ->cystic

PP ->INNP

EX ->There

VB->VBZNP

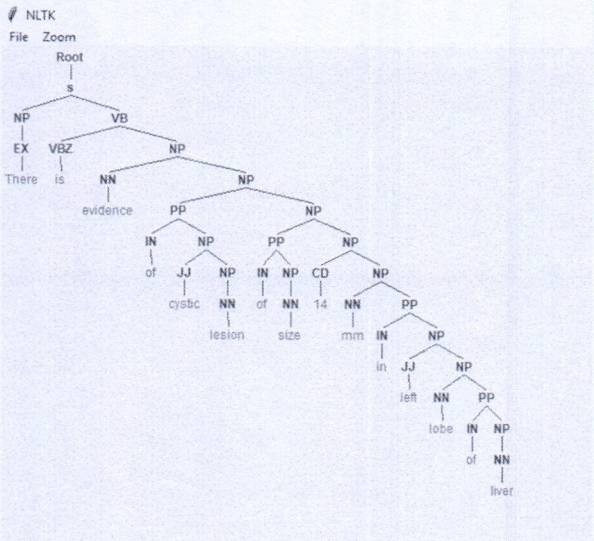
VBZ ->is

IN ->of|in

CD ->14

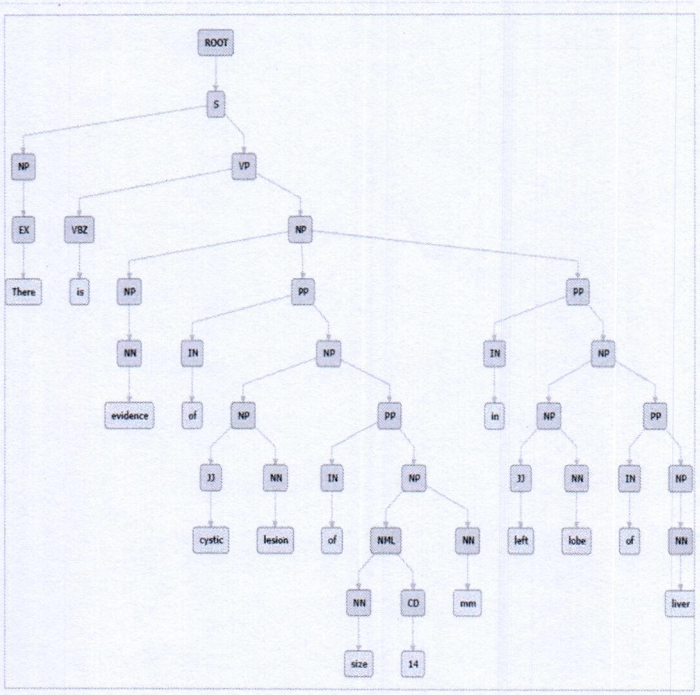
5.2.3.1 Result and Discussion

```
(Root
  (s
    (NP (EX There))
    (VB
      (VBZ is)
      (NP
        (NN evidence)
        (NP
          (PP (IN of) (NP (JJ cystic) (NP (NN lesion))))
          (NP
            (PP (IN of) (NP (NN size)))
            (NP
              (CD 14)
              (NP
                (NN mm)
                (PP
                  (IN in)
                  (NP
                    (JJ left)
                    (NP (NN lobe) (PP (IN of) (NP (NN liver))))))))))))))
['There', 'is', 'evidence', 'of', 'cystic', 'lesion', 'of', 'size', '14', 'mm', 'in', 'left', 'lobe', 'of', 'liver']
15
```



5.2.3.2 Compare with Stanford Parser

Here left is taken as adjective while in previous example it is taken as verb.



5.2.4 Example 4

Rule 1: adjective followed by noun phrase or preposition phrase followed by noun phrase or singular noun or cardinal number followed by noun phrase

Rule 2: preposition followed by noun phrase

S -> NP

NP ->JJNP|PPNP|NN|CDNP

JJ ->Enlarged

PP -> INNP

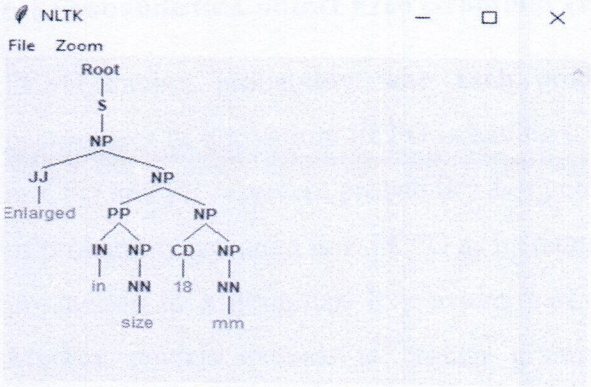
IN ->in

NN ->size | mm

CD ->18

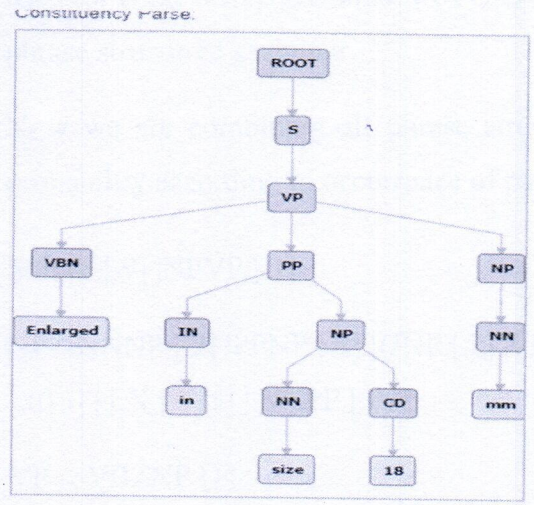
5.2.4.1 Result and Discussion

```
============= RESTART: C:\Users\hp\Desktop\phd presentation\ob4.py ===========
Grammar with 11 productions (start state = S)
    S -> NP
    NP -> JJNP
    NP -> PPNP
    NP -> NN
    NP -> CDNP
    JJ -> Enlarged
    PP -> INNP
    IN -> in
    NN -> size
    NN -> mm
    CD -> 18
(Root
  (S
    (NP
      (JJ Enlarged)
      (NP (PP (IN in) (NP (NN size))) (NP (CD 18) (NP (NN mm)))))))
['Enlarged', 'in', 'size', '18', 'mm']
8
```



5.2.4.2 Compare with Stanford Parser

Here enlarged is treated as a verb but according to sentence it is adjective



In phrase Structure Grammar ambiguity issue arises ,so to eliminate ambiguity we use probabilistic context free grammar .

Figure 6.3.2 Parse tree for "No evidence of lymphadenopathy in the neck region"

Bracket notation for Figure 6.3.2 is shown in Figure 6.3.3

(Root (S (NP (NP (DT No) (NP (NN evidence)))

(PP (IN of) (NP (NP (NN lymphadenopathy))

(PP (IN in) (NP (DT the) (NP (NN neck) (NP (NN region)))

)))))))

Figure 6.3.3 Bracket notation for "No evidence of lymphadenopathy in the neck region"

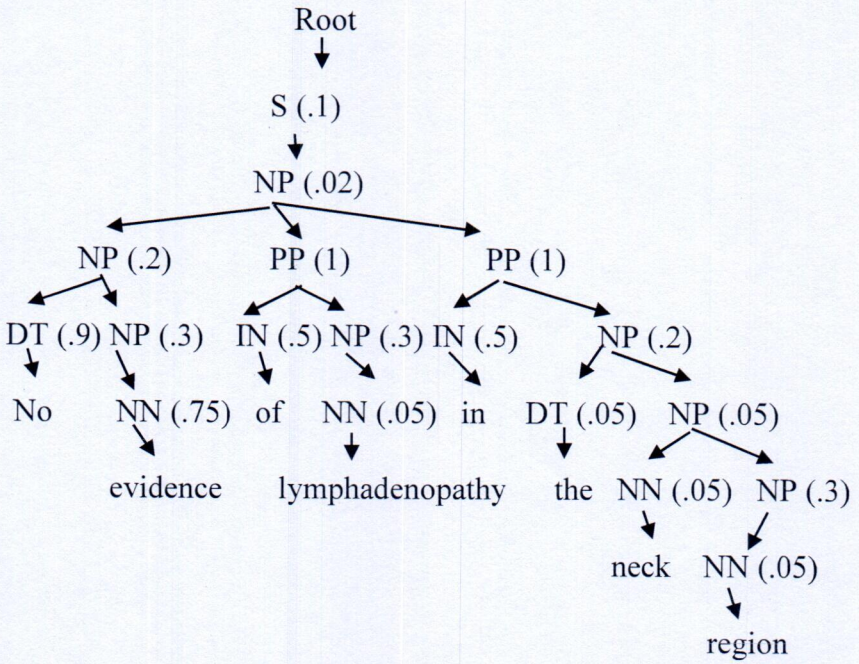


Figure 6.3.4. Parse Tree for "No evidence of lymphadenopathy in the neck region"

Bracket notation for Figure 6.3.4 is listed in Figure 6.3.5

(Root (S (NP (NP (DT No) (NP (NN evidence)))

(PP (IN of) (NP (NN lymphadenopathy)))

(PP (IN in) (NP (DT the) (NP (NN neck) (NP (NN region)))))

)))

Figure 6.3.5 Bracket notation for "No evidence of lymphadenopathy in the neck region"

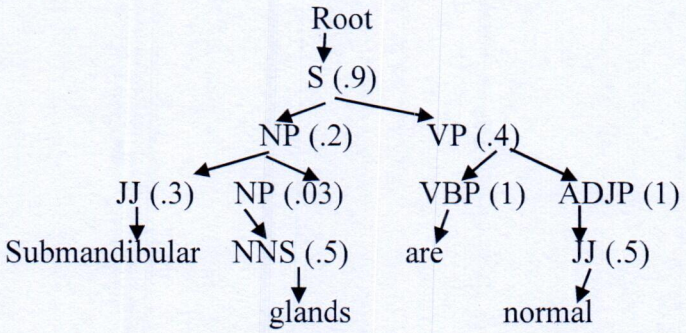


Figure 6.3.6. Parse Tree for "Submandibular glands are normal"

Bracket notation for Figure 6.3.6 is shown in Figure 6.3.7

(Root (S (NP (JJ submandibular) (NP (NNS glands)))

(VP (VBP are) (ADJP (JJ normal)))))

Figure 6.3.7 Bracket Notation for "Submandibular glands are normal"

Chapter 7

Analysis of other Reports

7.1USG of neck, cheek and lower abdomen

Grammar is different for each report. Each production is having probability according to occurrence in clinical report. If two production starting from same phrase, then phrase which is having high probability is prefer over less probability production. We select ultrasound reports of neck, cheek and chest reports of a hospital (name confidential). All reports are real reports of patients. For each sentence in a report tree will be generated. Grammar for cheek and neck reports is listed below.

S->NPVP .9/NP.1

NP->EX .05/NPPP .19/NN .05/NPVP .1/NPPPPP .05/NNCDNN .05/DTNP .05/NNS .05/NPCCNP .05/CDNN .005/DTJJRNN .005/JJNP .1 / NNNP .1/JJ .1/ ADJPNN .05

VP->VBZNP .5/VBNPP .1/VBPADJP .1/VBNP .1/VBZVP .1/VBN .1

IN->of .5/in .5

PP->INNP 1

EX->THERE 1

CD->6X2 .5/10X2 .3 /18X10 .2

DT->No .2/The .5/Both .3

CC->and 1

JJR->larger 1

ADJP->JJ .5/CDNN .4/ ADJPCCADJP .1

VBZ->is .5/shows .5

VBN->seen 1

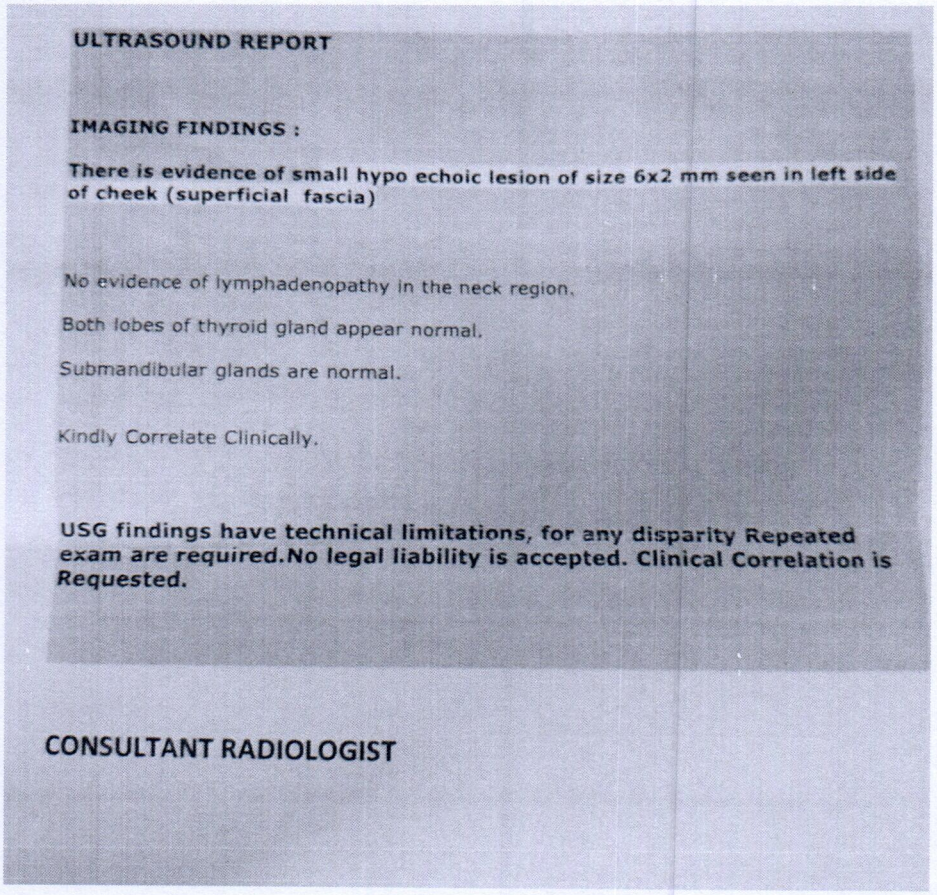
VBP-> appear .5/are .5

VB->define 1

NN->evidence .09/lenses .05/size .05/mm .05/side .05/cheek .05/facia .05 /lymphadenopathy .05/neck .05/region .05/thyroid .05/gland .05/ill .05/quadrant .05 /lesion .05/centre .05/s/o .005/collection .05/fluid .05/cavity .05/gluteal .003/USG .001/abnormality .001

NNS->lobes .5/glands .5

JJ->normal .3/small .05/hypo .05/echoic .05/super .05/left .05/right .05/outer .05/upper .05/obvious .05/focal .05/free .05/bilateral .05/submandibular .05/pleural .04/significant .01



ULTRASOUND REPORT

IMAGING FINDINGS :

There is evidence of small hypo echoic lesion of size 6x2 mm seen in left side of cheek (superficial fascia)

No evidence of lymphadenopathy in the neck region.

Both lobes of thyroid gland appear normal.

Submandibular glands are normal.

Kindly Correlate Clinically.

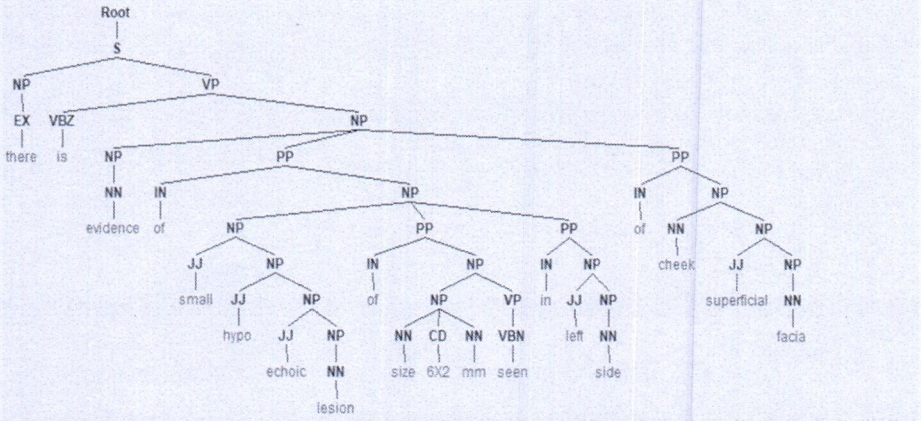
USG findings have technical limitations, for any disparity Repeated exam are required.No legal liability is accepted. Clinical Correlation is Requested.

CONSULTANT RADIOLOGIST

Figure 7.1 Ultrasound report of neck for patient 1

NLTK

File Zoom



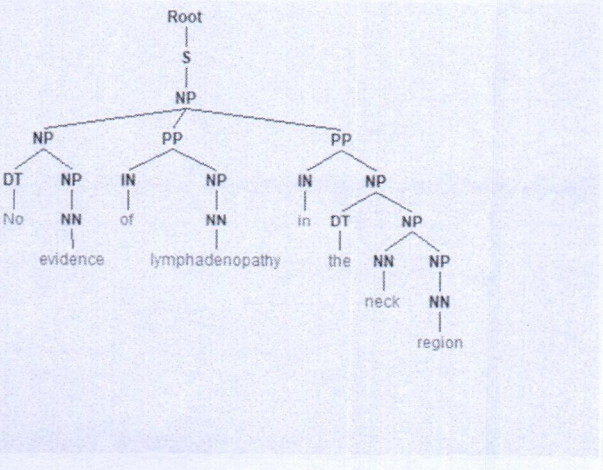

```
============ RESTART: C:\Users\hp\Desktop\phd presentation\final.py ============
```

Squeezed text (86 lines).

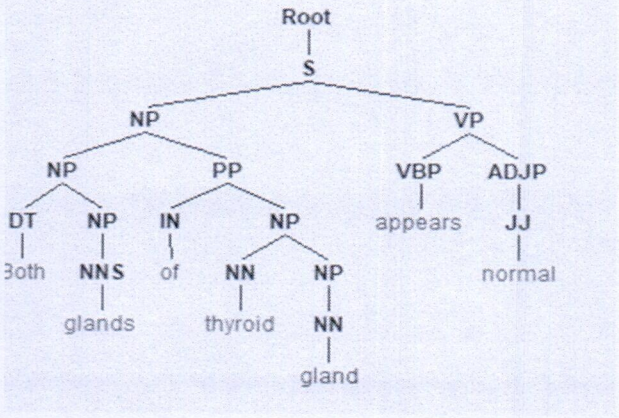
```
(Root
  (S
    (NP (EX there))
    (VP
      (VBZ is)
      (NP
        (NP (NN evidence))
        (PP
          (IN of)
          (NP
            (NP
              (JJ small)
              (NP (JJ hypo) (NP (JJ echoic) (NP (NN lesion)))))
            (PP
              (IN of)
              (NP (NP (NN size) (CD 6X2) (NN mm)) (VP (VBN seen))))
            (PP (IN in) (NP (JJ left) (NP (NN side))))))
        (PP
          (IN of)
          (NP (NN cheek) (NP (JJ superficial) (NP (NN facia))))))))))
['there', 'is', 'evidence', 'of', 'small', 'hypo', 'echoic', 'lesion', 'of', 'size', '6X2', 'mm', 'seen', 'in', 'left', 'side', 'of', 'cheek', 'superficial', 'facia']
12
8.7144970893855995e-36
```

NLTK

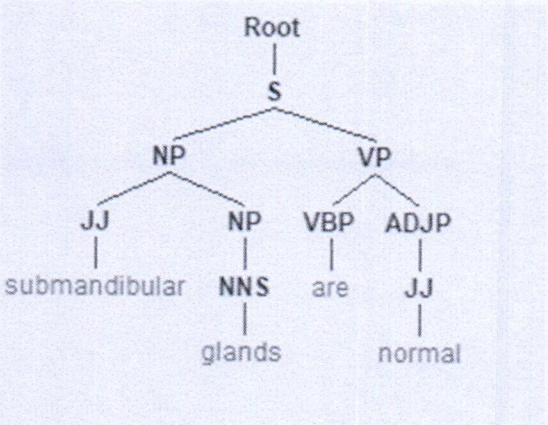
File Zoom



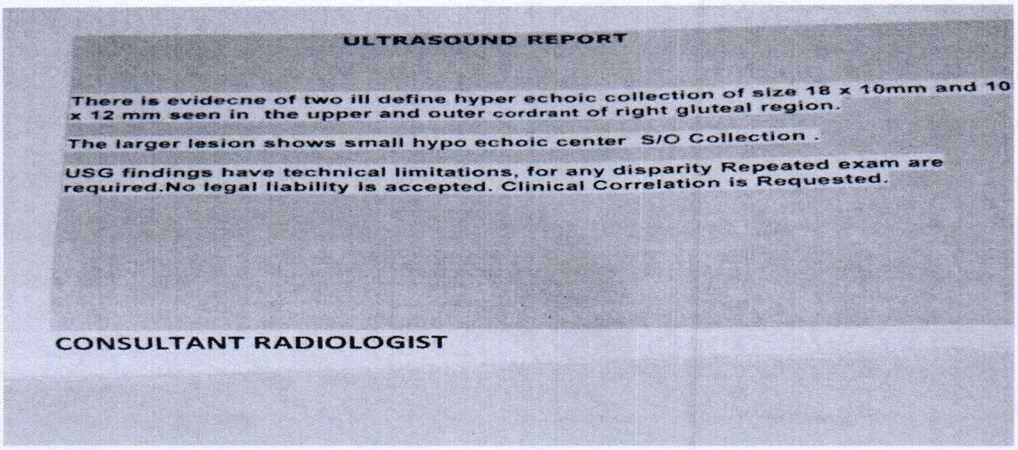
```
(Root
  (NP
    (NP (DT No) (NP (NN evidence)))
    (PP (IN of) (NP (NN lymphadenopathy)))
    (PP (IN in) (NP (DT the) (NP (NN neck) (NP (NN region)))))))
['No', 'evidence', 'of', 'lymphadenopathy', 'in', 'the', 'neck', 'region']
8
9.277343750000009e-17
```




```
(Root
  (S
    (NP
      (NP (DT Both) (NP (NNS glands)))
      (PP (IN of) (NP (NN thyroid) (NP (NN gland)))))
    (VP (VBP appears) (ADJP (JJ normal)))))
['Both', 'glands', 'of', 'thyroid', 'gland', 'appears', 'normal']
8
7.910156250000003e-13
```




```
(Root
  (S
    (NP (JJ submandibular) (NP (NNS glands)))
    (VP (VBP are) (ADJP (JJ normal)))))
['submandibular', 'glands', 'are', 'normal']
6
5.0625e-06
```



ULTRASOUND REPORT

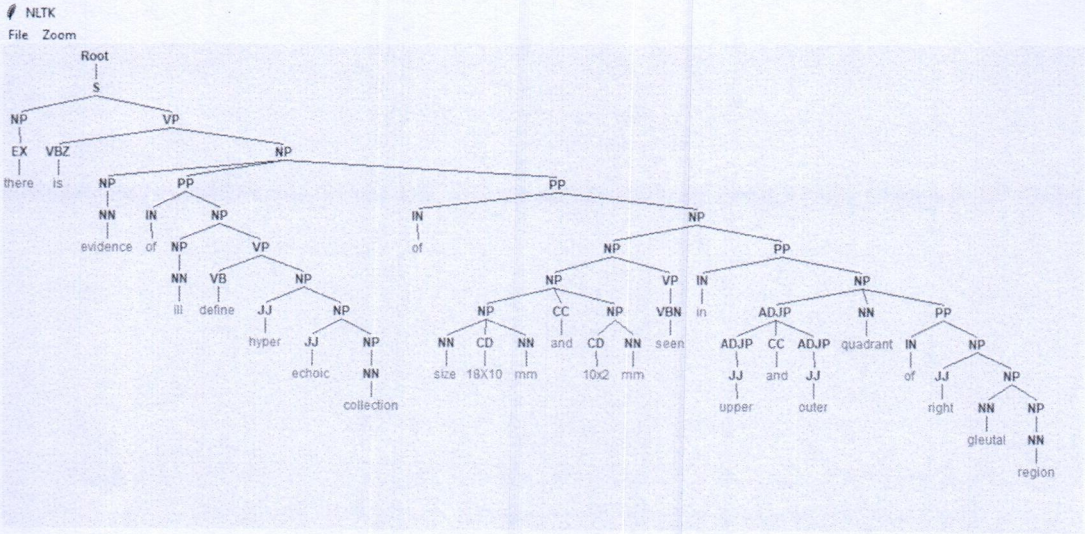
There is evidecne of two ill define hyper echoic collection of size 18 x 10mm and 10 x 12 mm seen in the upper and outer cordrant of right gluteal region.

The larger lesion shows small hypo echoic center S/O Collection .

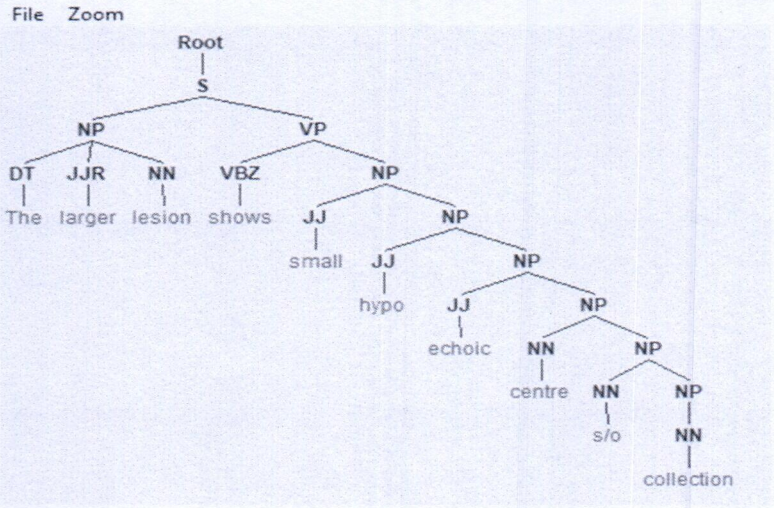
USG findings have technical limitations, for any disparity Repeated exam are required.No legal liability is accepted. Clinical Correlation is Requested.

CONSULTANT RADIOLOGIST

Figure 7.1.1. Ultrasound report of abdomen




```
(Root
  (S
    (NP (EX there))
    (VP
      (VBZ is)
      (NP
        (NP (NN evidence))
        (PP
          (IN of)
          (NP
            (NP (NN ill))
            (VP
              (VB define)
              (NP (JJ hyper) (NP (JJ echoic) (NP (NN collection)))))))
        (PP
          (IN of)
          (NP
            (NP
              (NP
                (NP (NN size) (CD 18X10) (NN mm))
                (CC and)
                (NP (CD 10x2) (NN mm)))
              (VP (VBN seen)))
            (PP
              (IN in)
              (NP
                (ADJP (ADJP (JJ upper)) (CC and) (ADJP (JJ outer)))
                (NN quadrant)
                (PP
                  (IN of)
                  (NP (JJ right) (NP (NN gleutal) (NP (NN region))))))))))))
['there', 'is', 'evidence', 'of', 'ill', 'define', 'hyper', 'echoic', 'collection', 'of', 'size', '18X10', 'mm', 'and', '10x2', 'mm', 'seen', 'in', 'upper', 'and', 'ou
ter', 'quadrant', 'of', 'right', 'gleutal', 'region']
14
3.7254475057125135e-47
```




```
(Root
  (S
    (NP (DT The) (JJR larger) (NN lesion))
    (VP
      (VBZ shows)
      (NP
        (JJ small)
        (NP
          (JJ hypo)
          (NP
            (JJ echoic)
            (NP (NN centre) (NP (NN s/o) (NP (NN collection)))))))))
['The', 'larger', 'lesion', 'shows', 'small', 'hypo', 'echoic', 'centre', 's/o', 'collection']
11
2.1972656250000017e-20
```

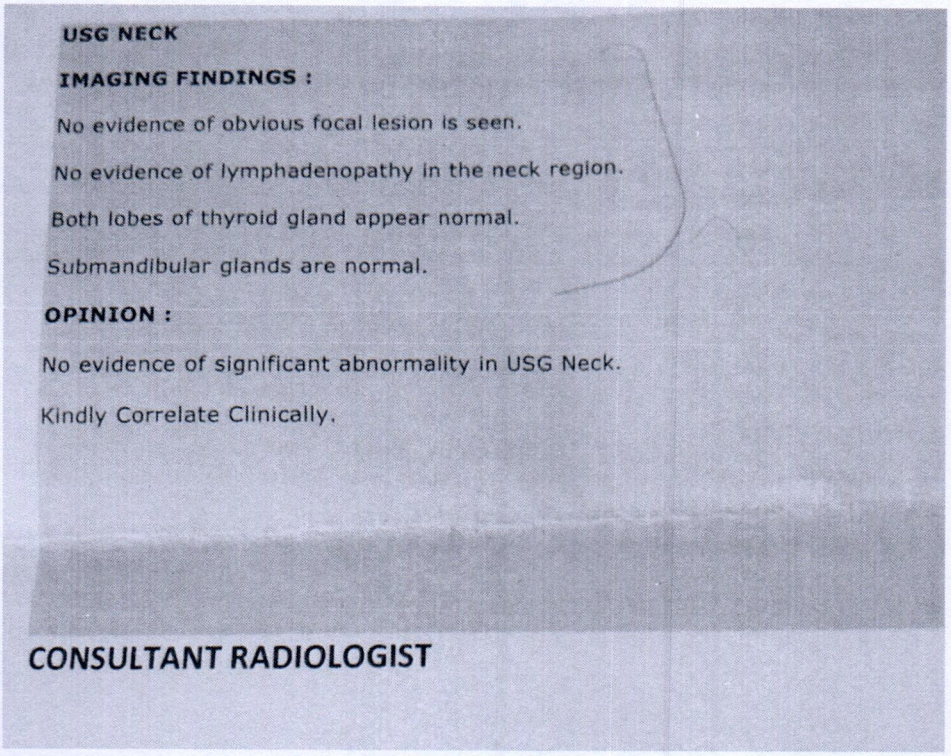
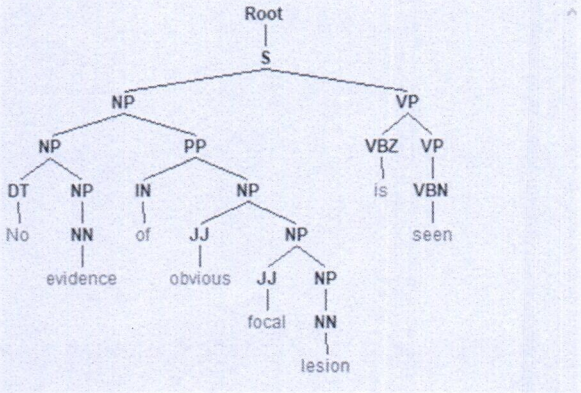

USG NECK
IMAGING FINDINGS :
No evidence of obvious focal lesion is seen.
No evidence of lymphadenopathy in the neck region.
Both lobes of thyroid gland appear normal.
Submandibular glands are normal.
OPINION :
No evidence of significant abnormality in USG Neck.
Kindly Correlate Clinically.
CONSULTANT RADIOLOGIST

Figure 7.1.3 Ultrasound report of neck for patient 2

NLTK

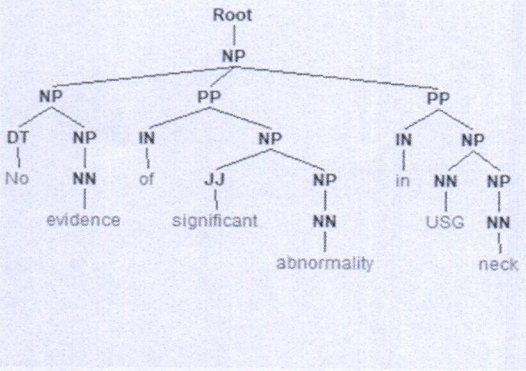
File Zoom




```
(Root
  (S
    (NP
      (NP (DT No) (NP (NN evidence)))
      (PP (IN of) (NP (JJ obvious) (NP (JJ focal) (NP (NN lesion))))))
    (VP (VBZ is) (VP (VBN seen)))))
['No', 'evidence', 'of', 'obvious', 'focal', 'lesion', 'is', 'seen']
9
1.5820312500000015e-17
```

NLTK

File Zoom




```
(Root
  (NP
    (NP (DT No) (NP (NN evidence)))
    (PP (IN of) (NP (JJ significant) (NP (NN abnormality))))
    (PP (IN in) (NP (NN USG) (NP (NN neck))))))
['No', 'evidence', 'of', 'significant', 'abnormality', 'in', 'USG', 'neck']
7
5.076562500000003e-21
```

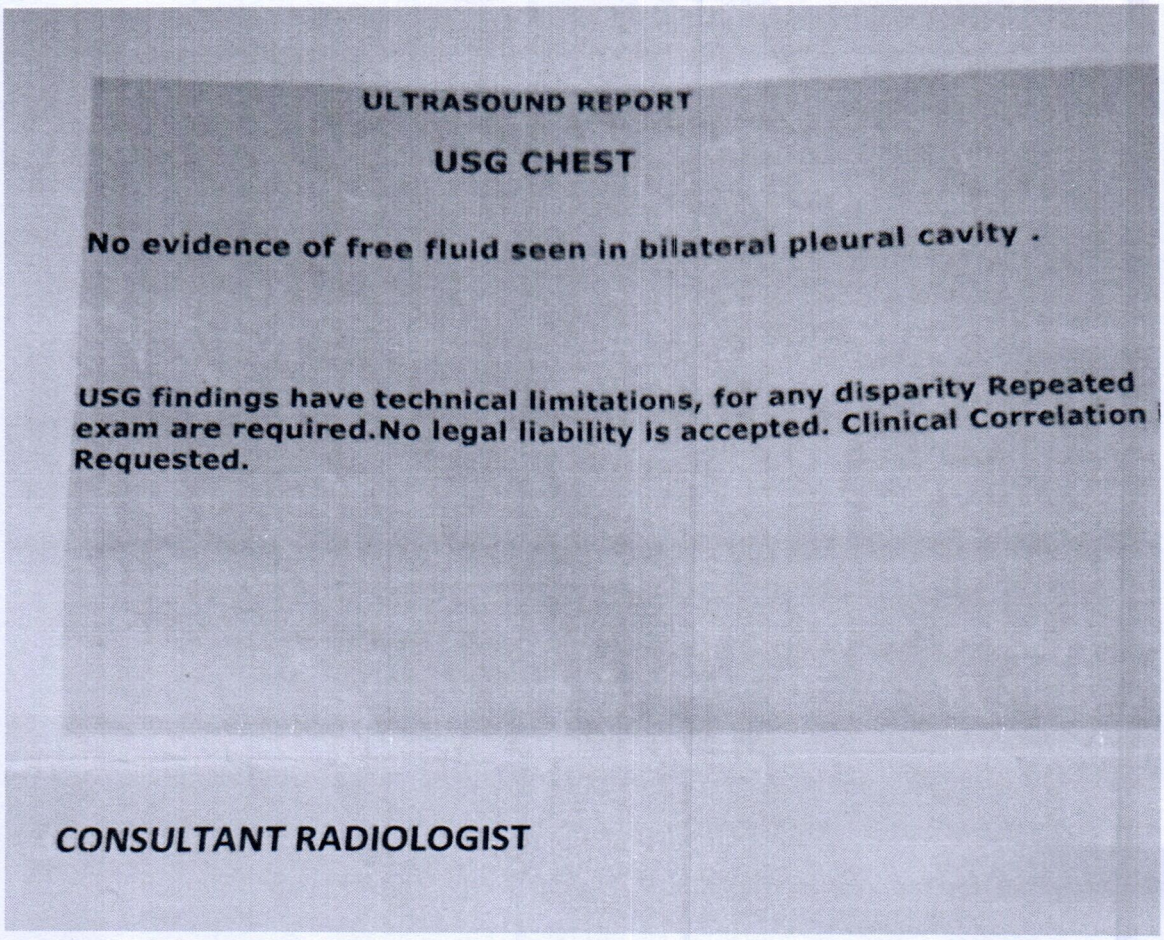
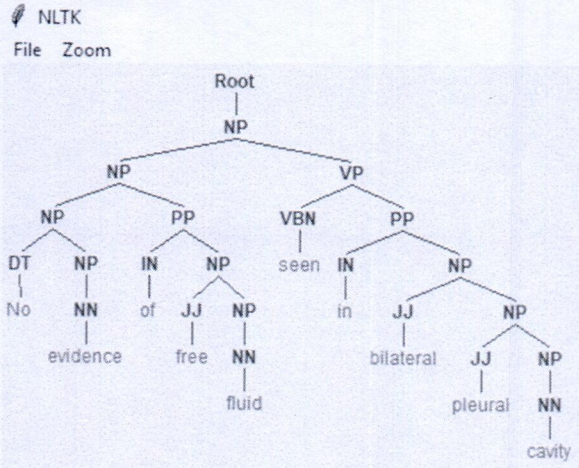

ULTRASOUND REPORT
USG CHEST
No evidence of free fluid seen in bilateral pleural cavity .
USG findings have technical limitations, for any disparity Repeated exam are required.No legal liability is accepted. Clinical Correlation Requested.
CONSULTANT RADIOLOGIST

Figure 7.1.4 Ultrasound report of chest




```
(Root
  (NP
    (NP
      (NP (DT No) (NP (NN evidence)))
      (PP (IN of) (NP (JJ free) (NP (NN fluid)))))
    (VP
      (VBN seen)
      (PP
        (IN in)
        (NP (JJ bilateral) (NP (JJ pleural) (NP (NN cavity))))))))
['No', 'evidence', 'of', 'free', 'fluid', 'seen', 'in', 'bilateral', 'pleural', 'cavity']
9
5.343750000000005e-22
```

7.2 USG report

S -> FRAG [1]

NP -> NP VP [.05] | NN [.1] | NP CC NP [.1] | DT [.05] | VBN NN [.05] | NP PRN [.05] | JJ NP .[1] | NP NP [.1] | NP , NP CC NP [.05] | VBD NP [.05] | EX [.03] | NP PP [.05] | NNS [.05] | NN CD NN [.1] | NN / JJ NNS[.03] |ADVP [.015] | ADJP [.02] | NP / NP[.005]

VP -> VBD [.1] | VBG NP [.1] | VBP ADJP [.1] | VBP ADJP PP [.1] | VBZ VP [.1] | VBN [.1] | RB VP [.05] |VBZ NP [.1] | VBN PP [.1] | VBG: NP VP [.05] | VBZ ADJP [.1]

PP -> IN NP [1]

IN->in [.5] | at [.2] | of [.3]

JJ -> normal [.1] |pancreatic [.05] | focal [.05] | medullary [.05] | echogenic [.05] | enlarged [.05] | free [.05] | mild [.05] | paraaortic [.05] | obvious [.05] | internal[.05] |soft [.05] | upper [.05] | peritorial [.05] | thickened [.05] | irregular [.05] | endometrial [.05] | distended [.1]

NN -> size [.1] | shape [.1] | echotexture [.1] | calibre [.05] | course [.05] | wall [.05] | thickness [.05] | evidence [.05] | kidney [.005] | e/o [.005] | inter [.005] | lumia [.005] | mass [.05] | lesion [.05] | duct [.005] | splenvein [.005] | IHBR [.005] | Cartico [.005] | differentiation [.005] | dilation [.005] |collection [.005] | gb [.003] | lumen [.002] |mm .005 | calculi [.005] | lymphadenopathy[.005] | cavity [.005] | fluid [.005] | abdomen [.005] | pelvis [.005] | calculus [.003] | vol .[002] | cc [.003] | sludged [.005] | ball[.005] | pelvicalyceal [.002] | system [.002] | foci [.005] |concretion [.003] | pole[.003] | region [.005] | bladder [.002] | liver [.05] |PV [.05] |CBD [.05]|GB [.005] |uterus [.005]| prostate [.005]|spleen [.005] | urinary [.002] |Impression[.043]

CC -> and [1]

VBD -> noted [.5] | raised [.5]

ADJP - > JJ PP [.1] | ADJP CC VP [.2] | JJ [.5] | ADJP CC ADJP [.1] |RB JJ [.1]

VBG -> showing [.3] | causting [.2] | thickening [.5]

DT -> no [.5] | both [.2]| the [.3]

VBP -> are [.7] | appear [.3]

VBN -> seen [.3] | left [.1] | maintained [.3] |dilated [.3]

VBZ -> is [.7] | shows [.3]

RB -> not [.3] | mildly [.1] | partially [.3] | well [.3]

ADVP -> JJ PP [1]

NNS -> pas [.05] | echoes [.05] |ovaries [.3] | pancreas [.3] | kidneys [.3]

PRN -> LRB NP RRB [1]

EX -> there [1]

CD -> -8 .5[.5] | -36 [.5]

: ->: [1]

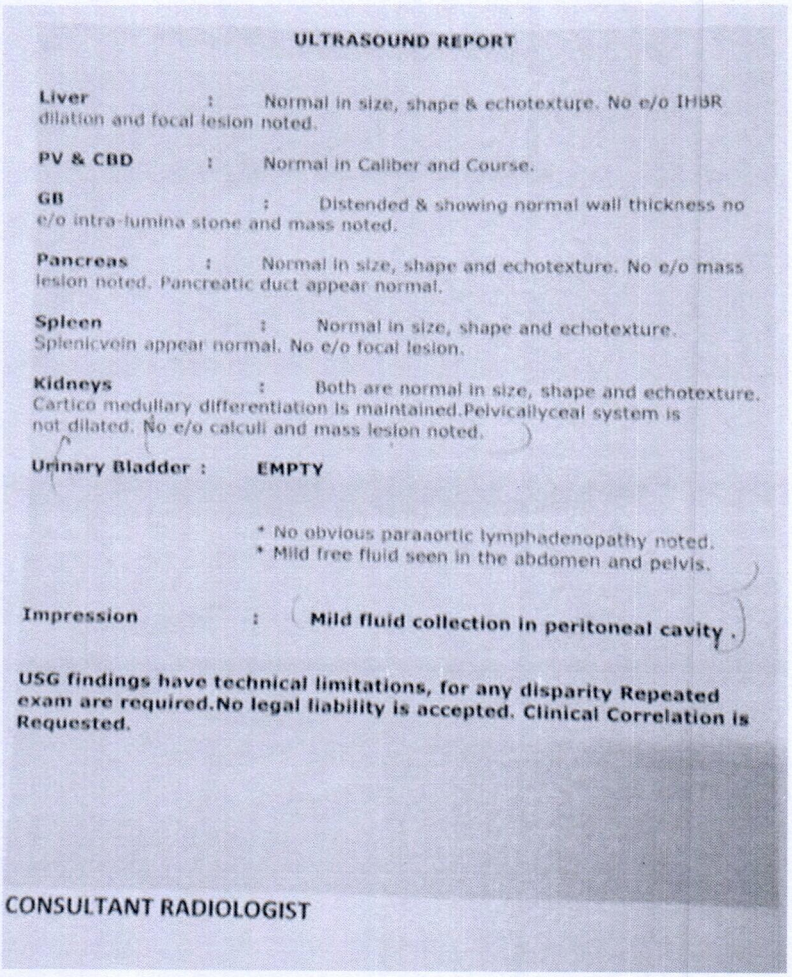
, ->, [1]

/ -> / [1]

LRB -> ([1]

RRB ->) [1]

FRAG -> NP: NP [.8] | NP VP [.2]



ULTRASOUND REPORT

Liver : Normal in size, shape & echotexture. No e/o IHBR dilation and focal lesion noted.

PV & CBD : Normal in Caliber and Course.

GB : Distended & showing normal wall thickness no e/o intra-lumina stone and mass noted.

Pancreas : Normal in size, shape and echotexture. No e/o mass lesion noted. Pancreatic duct appear normal.

Spleen : Normal in size, shape and echotexture. Splenicvein appear normal. No e/o focal lesion.

Kidneys : Both are normal in size, shape and echotexture. Cartico medullary differentiation is maintained.Pelvicailyceal system is not dilated. No e/o calculi and mass lesion noted.

Urinary Bladder : EMPTY

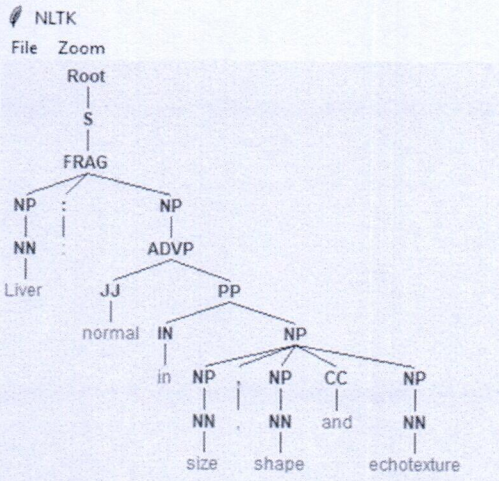
* No obvious paraaortic lymphadenopathy noted.
* Mild free fluid seen in the abdomen and pelvis.

Impression : Mild fluid collection in peritoneal cavity .

USG findings have technical limitations, for any disparity Repeated exam are required.No legal liability is accepted. Clinical Correlation is Requested.

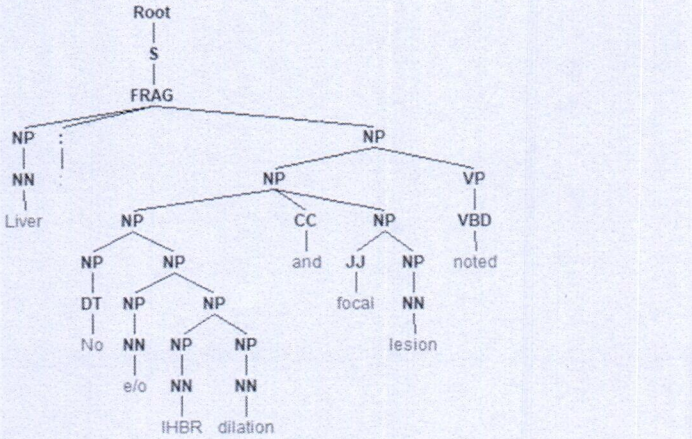
CONSULTANT RADIOLOGIST

Figure 7.2.1 Ultrasound Report 1



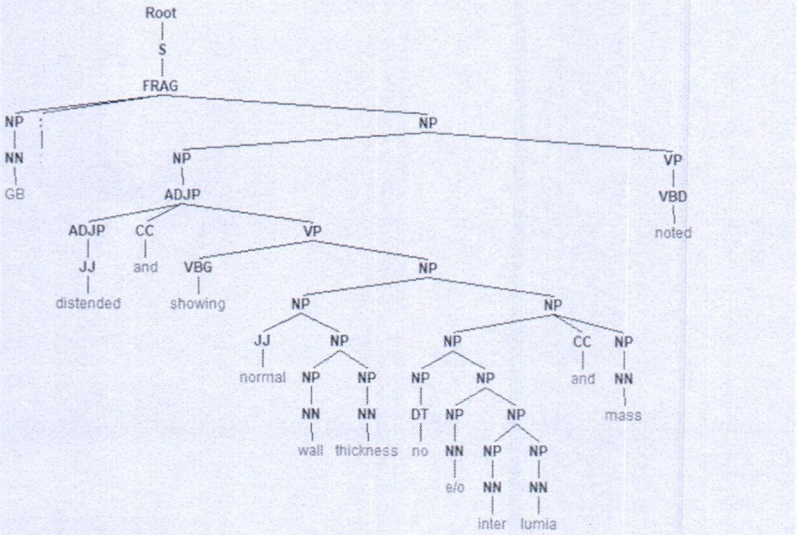
NLTK

File Zoom



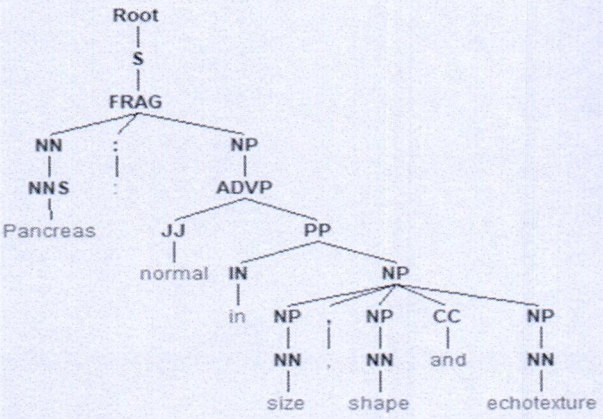
NLTK

File Zoom



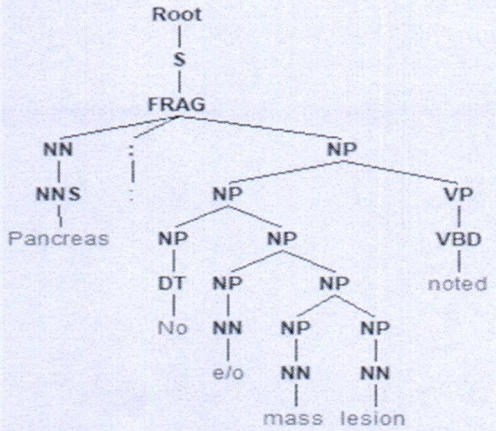
NLTK

File Zoom



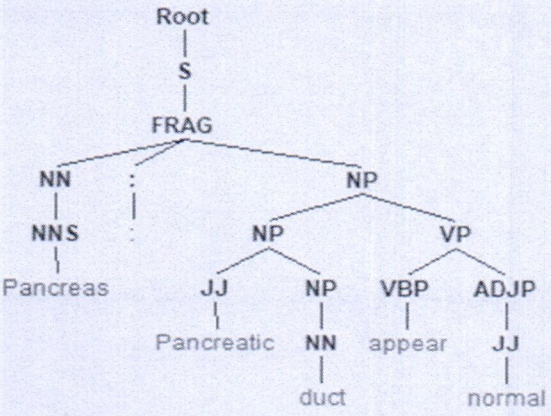
NLTK

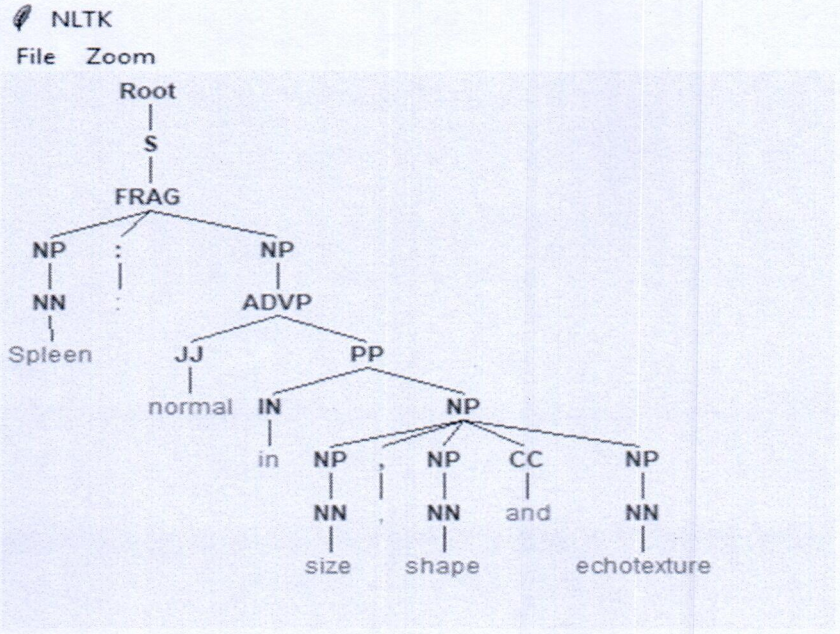
File Zoom

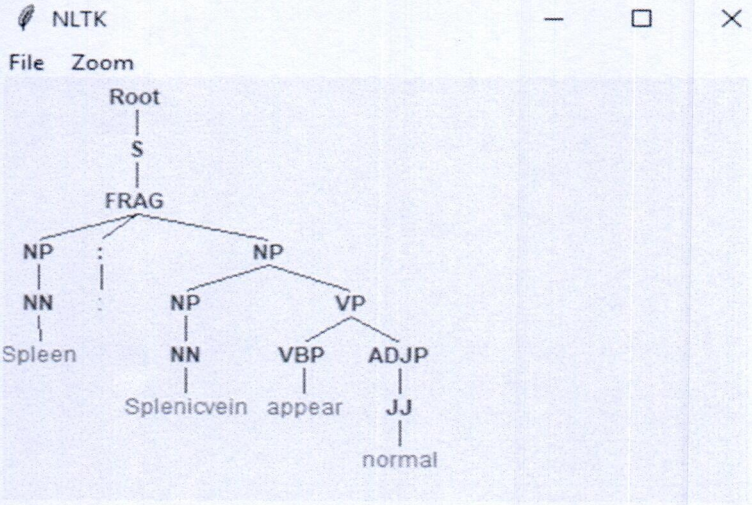


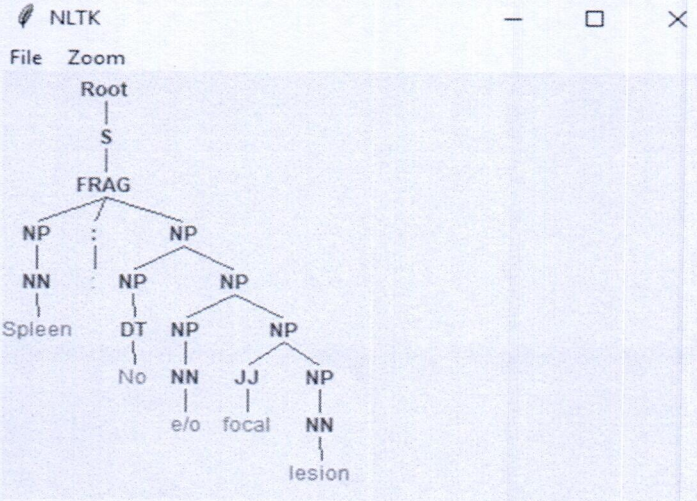
NLTK

File Zoom



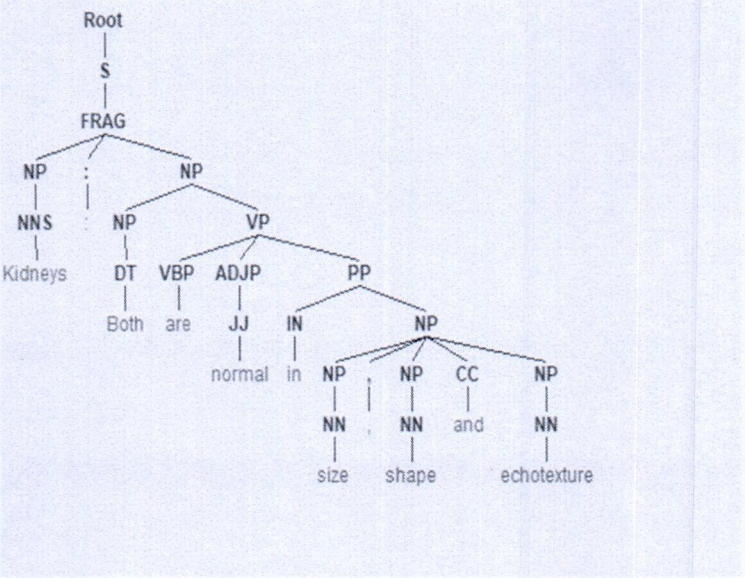

NLTK
File Zoom
Root
S
FRAG
NP
:
NP
NN
:
ADVP
Spleen
JJ
PP
normal
IN
NP
in
NP
,
NP
CC
NP
NN
,
NN
and
NN
size
shape
echotexture


NLTK
File Zoom
Root
S
FRAG
NP
:
NP
NN
:
NP
VP
Spleen
NN
VBP
ADJP
Splenicvein
appear
JJ
normal


NLTK
File Zoom
Root
S
FRAG
NP
:
NP
NN
:
NP
NP
Spleen
DT
NP
NP
No
NN
JJ
NP
e/o
focal
NN
lesion

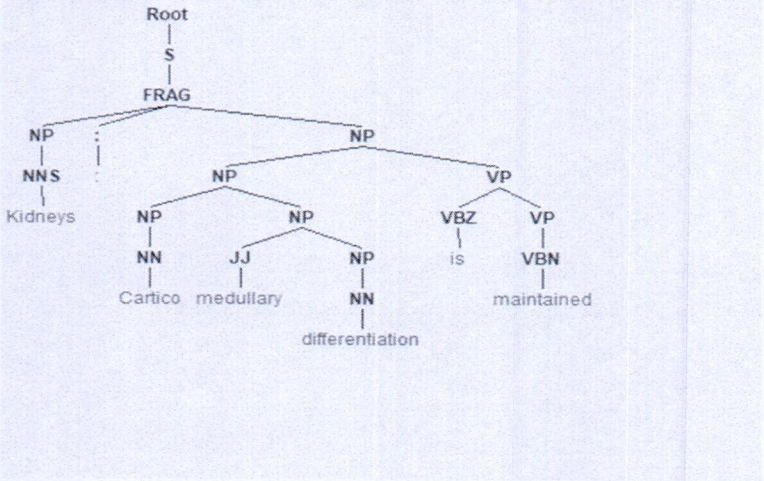
NLTK

File Zoom



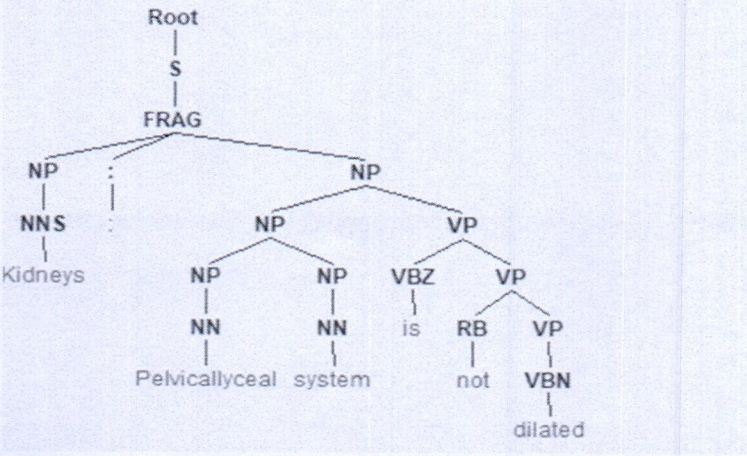
NLTK

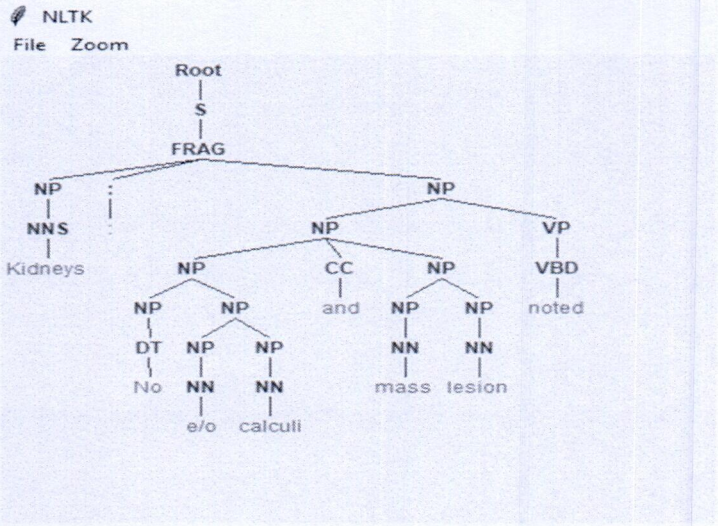
File Zoom

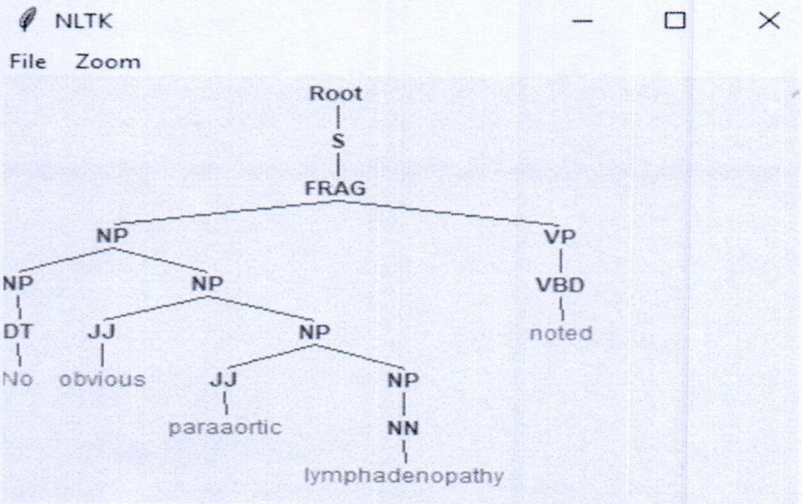


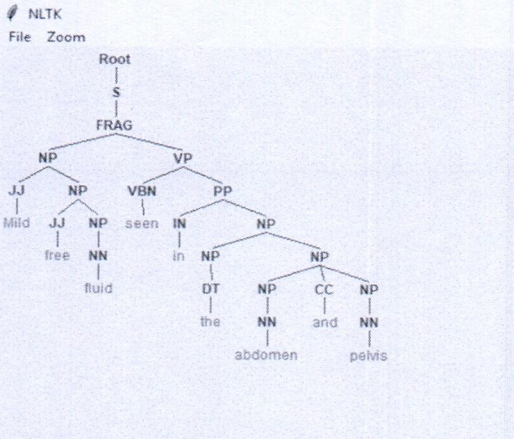
NLTK

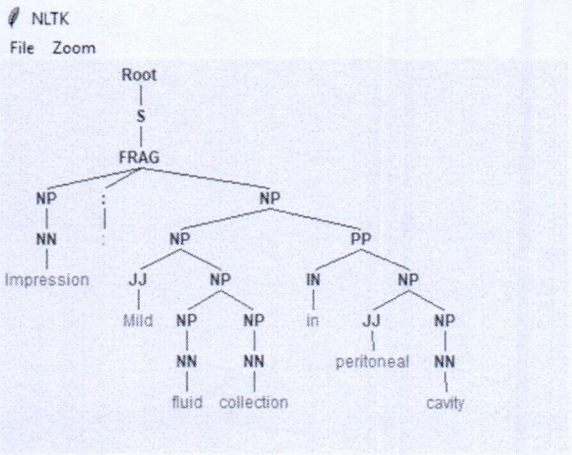
File Zoom




NLTK
File Zoom
Root
S
FRAG
NP
:
NP
NNS
NP
VP
Kidneys
NP
CC
NP
VBD
NP
NP
NP
NP
and
noted
DT
NP
NP
NN
NN
No
NN
NN
mass
lesion
e/o
calculi


NLTK
File Zoom
Root
S
FRAG
NP
VP
NP
NP
VBD
DT
JJ
NP
noted
No
obvious
JJ
NP
paraaortic
NN
lymphadenopathy


NLTK
File Zoom
Root
S
FRAG
NP
VP
JJ
NP
VBN
PP
Mild
JJ
NP
seen
IN
NP
free
NN
in
NP
NP
fluid
DT
NP
CC
NP
the
NN
and
NN
abdomen
pelvis



ULTRASOUND REPORT

Liver : Normal in size, shape & echotexture. No e/o IHBR dilation and focal lesion noted.

PV & CBD : Normal in Caliber and Course.

GB : Distended & showing normal wall thickness **There is evidence of echogenic foci not causting pas seen in gb lumen (size - 8mm)**

Pancreas : Normal in size, shape and echotexture. No e/o mass lesion noted. Pancreatic duct appear normal.

Spleen : Normal in size, shape and echotexture. Splenicvein appear normal. No e/o focal lesion.

Kidneys : Both are normal in size, shape and echotexture. Cartico medullary differentiation is maintained.Pelvicallyceal system is not dilated. No e/o calculi and mass lesion noted.

Urinary Bladder : Well distended. No e/o wal thickening/Calculi/internal echoes noted.

Uterus : Normal in size, shape and echotexture.No e/o mass lesion noted. Endometrial echoes appear normal.

Ovaries : Normal in size, shape and echotexture.No e/o mass lesion noted.

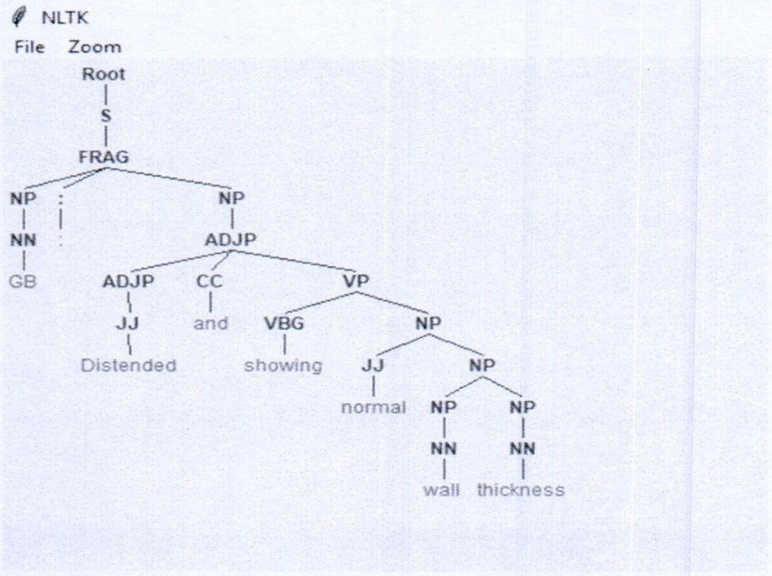
* No obvious paraaortic lymphadenopathy noted.
* No free fluid seen in the abdomen and pelvis.

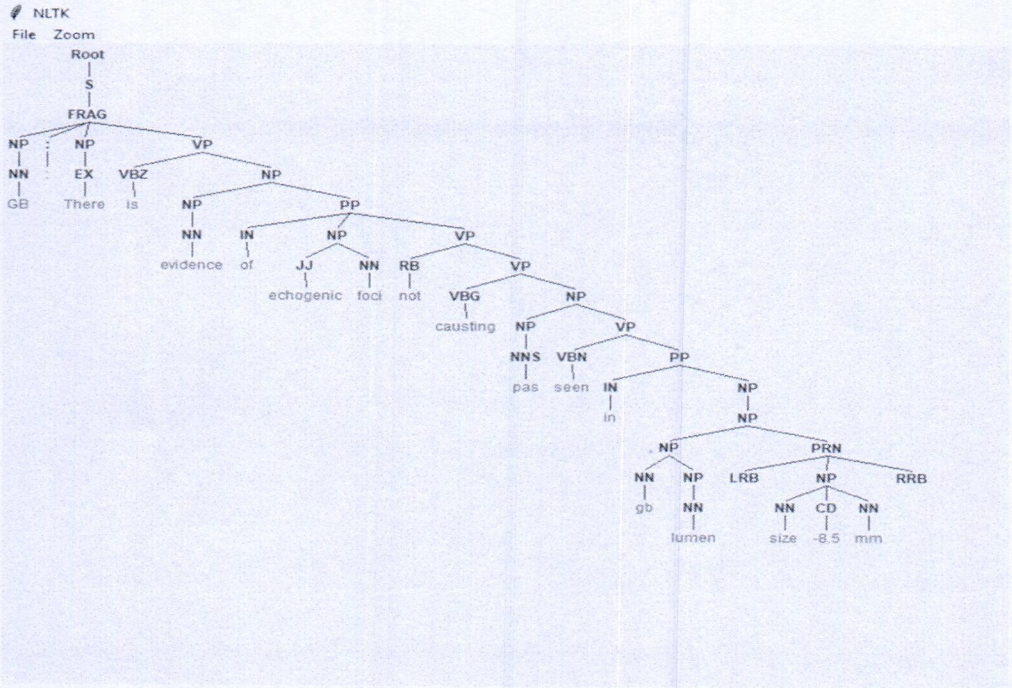
Impression : **Sludge ball /soft calculus in gb.**

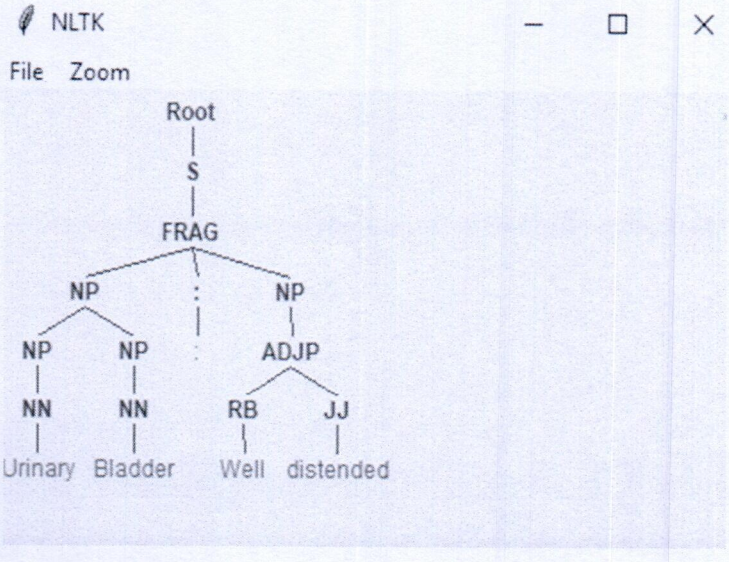
USG findings have technical limitations, for any disparity Repeated exam are required.No legal liability is accepted. Clinical Correlation is Requested.

CONSULTANT RADIOLOGIST

Figure 7.2.2 Ultrasound Report 2

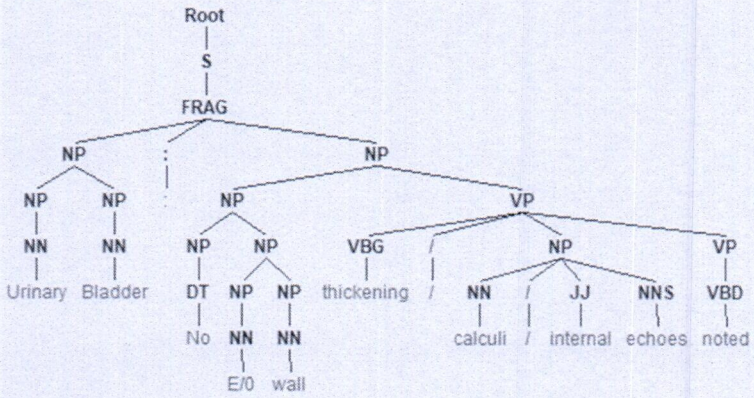

NLTK
File Zoom
Root
S
FRAG
NP : NP
NN : ADJP
GB ADJP CC VP
JJ and VBG NP
Distended showing JJ NP
normal NP NP
NN NN
wall thickness


NLTK
File Zoom
Root
S
FRAG
NP : NP VP
NN : EX VBZ NP
GB There Is NP PP
NN IN NP VP
evidence of JJ NN RB VP
echogenic foci not VBG NP
causting NP VP
NNS VBN PP
pas seen IN NP
in NP
NP PRN
NN NP LRB NP RRB
gb NN NN CD NN
lumen size -8.5 mm


NLTK
File Zoom
Root
S
FRAG
NP : NP
NP NP : ADJP
NN NN RB JJ
Urinary Bladder Well distended

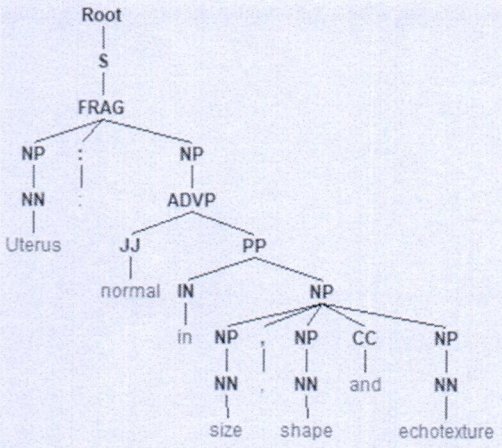
NLTK

File Zoom



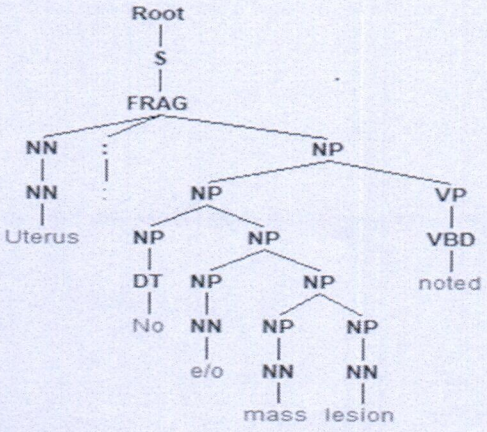
NLTK

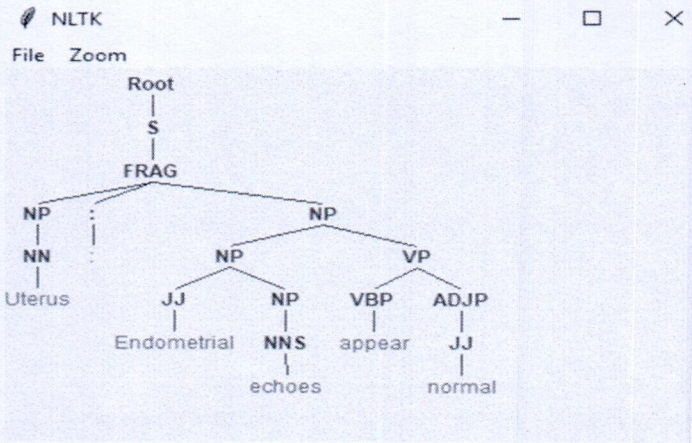
File Zoom

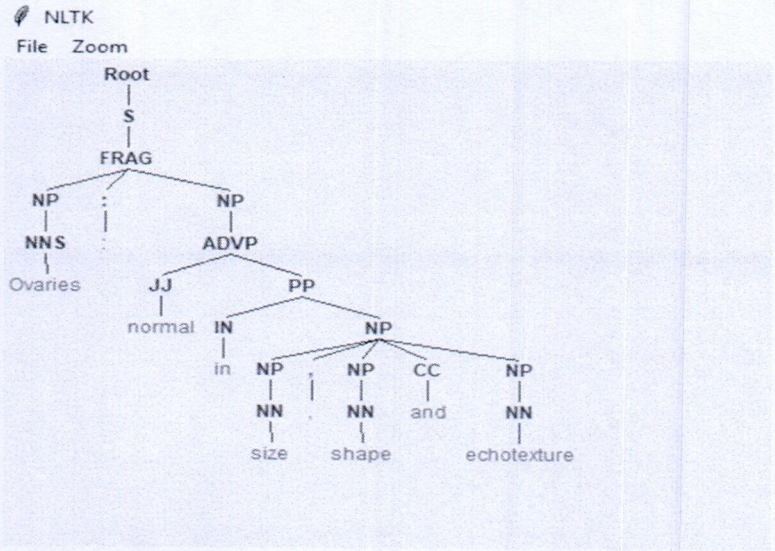


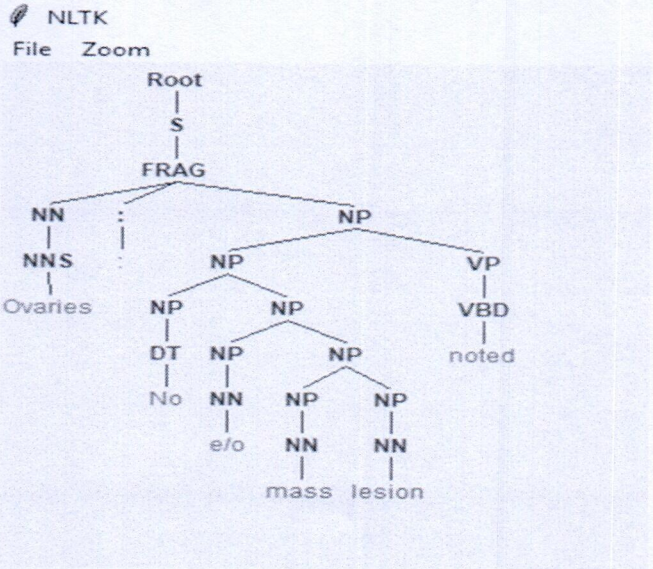
NLTK

File Zoom



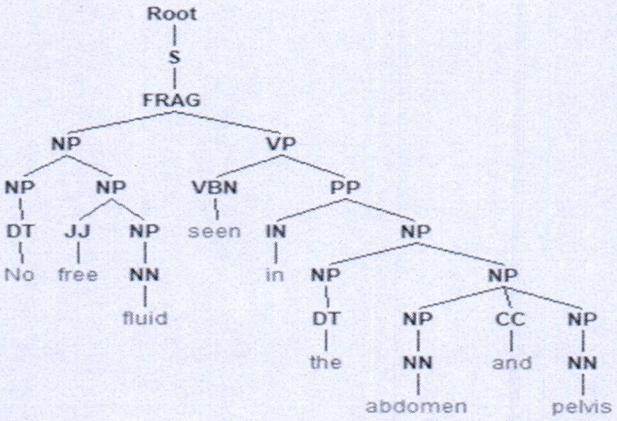

NLTK
File Zoom
Root
S
FRAG
NP
:
NP
NN
:
NP
VP
Uterus
JJ
NP
VBP
ADJP
Endometrial
NNS
appear
JJ
echoes
normal


NLTK
File Zoom
Root
S
FRAG
NP
:
NP
NNS
:
ADVP
Ovaries
JJ
PP
normal
IN
NP
in
NP
,
NP
CC
NP
NN
,
NN
and
NN
size
shape
echotexture


NLTK
File Zoom
Root
S
FRAG
NN
:
NP
NNS
:
NP
VP
Ovaries
NP
NP
VBD
DT
NP
NP
noted
No
NN
NP
NP
e/o
NN
NN
mass
lesion

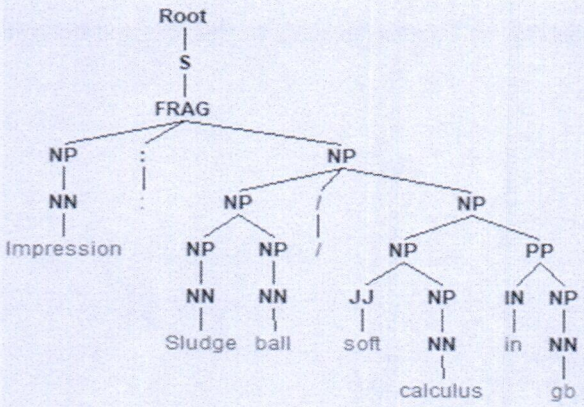
NLTK

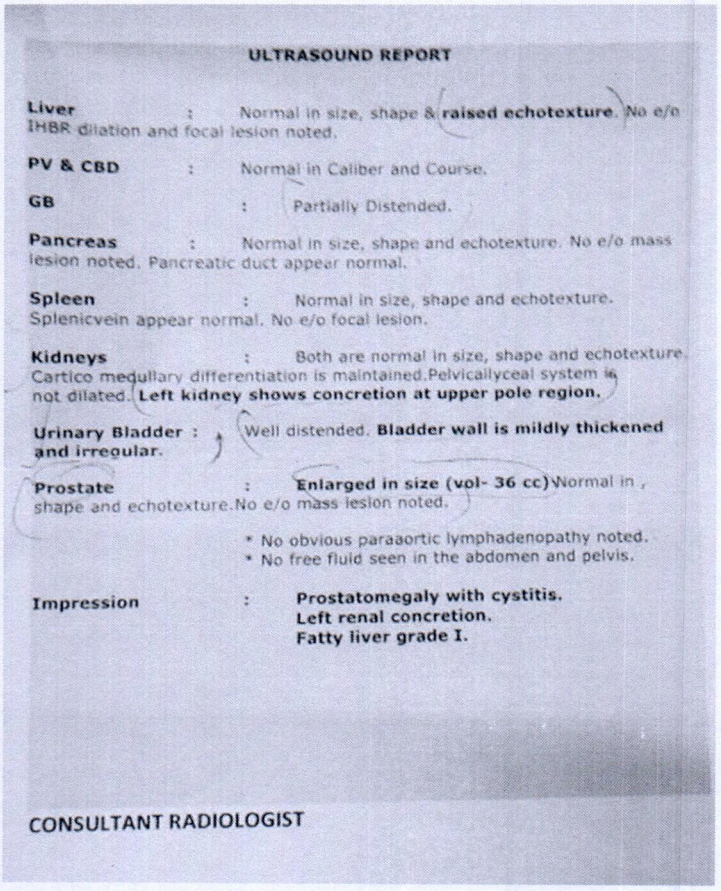
File Zoom



NLTK

File Zoom





ULTRASOUND REPORT

Liver : Normal in size, shape & **raised echotexture**. No e/o IHBR dilation and focal lesion noted.

PV & CBD : Normal in Caliber and Course.

GB : Partially Distended.

Pancreas : Normal in size, shape and echotexture. No e/o mass lesion noted. Pancreatic duct appear normal.

Spleen : Normal in size, shape and echotexture. Splenicvein appear normal. No e/o focal lesion.

Kidneys : Both are normal in size, shape and echotexture. Cartico medullary differentiation is maintained.Pelvicailyceal system is not dilated. **Left kidney shows concretion at upper pole region.**

Urinary Bladder : and irregular. Well distended. **Bladder wall is mildly thickened**

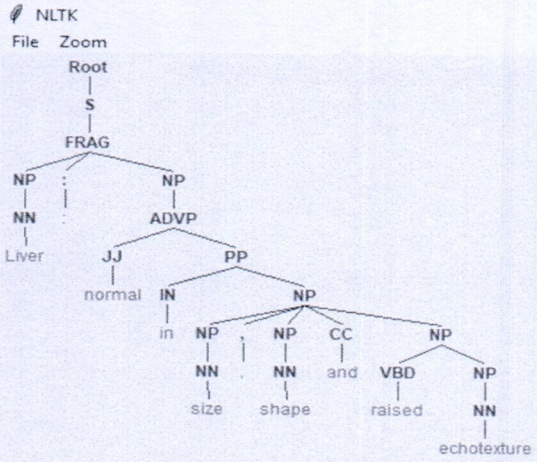
Prostate : **Enlarged in size (vol- 36 cc)** Normal in , shape and echotexture.No e/o mass lesion noted.

* No obvious paraaortic lymphadenopathy noted.
* No free fluid seen in the abdomen and pelvis.

Impression : **Prostatomegaly with cystitis.**
Left renal concretion.
Fatty liver grade I.

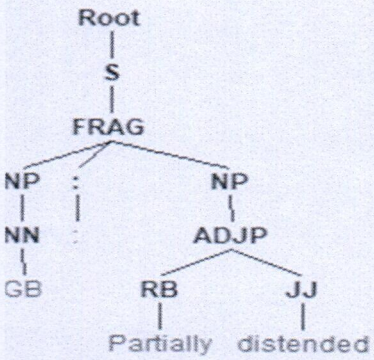
CONSULTANT RADIOLOGIST

Figure 7.2.3 Ultrasound report 3



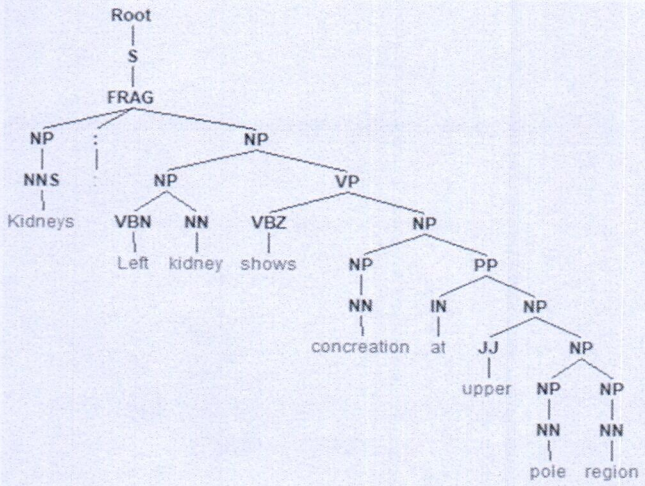
NLTK

File Zoom



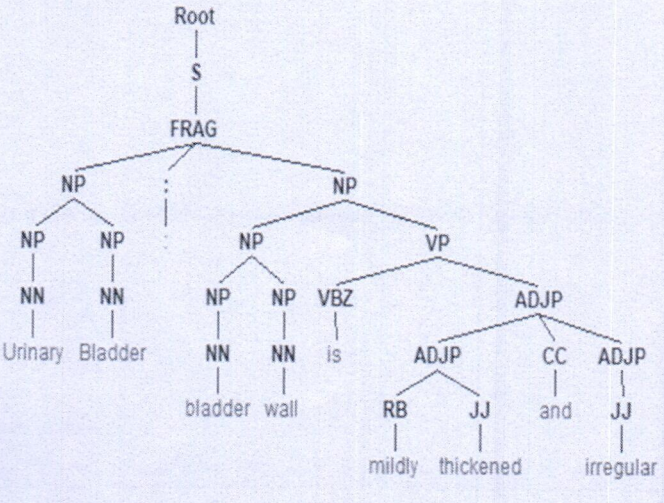
NLTK

File Zoom



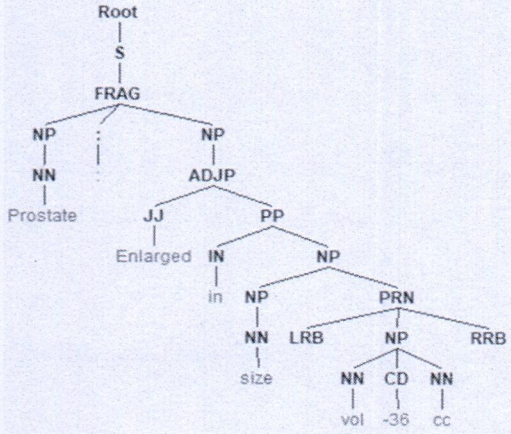
NLTK

File Zoom



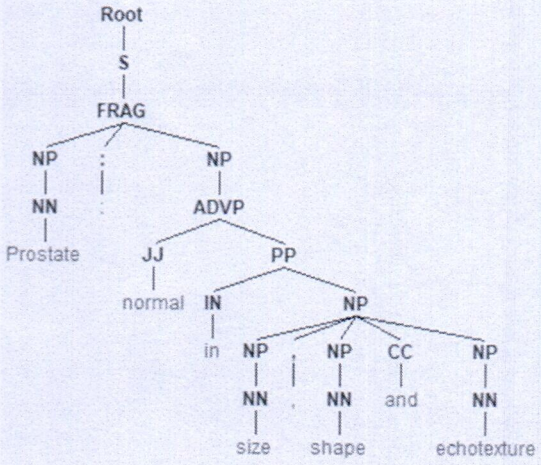
NLTK

File Zoom



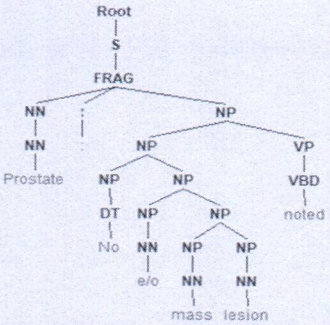
NLTK

File Zoom



NLTK

File Zoom



Conclusion and Future Scope

- Parser is generated with the help of NLP techniques as it is helpful in medical domain also. Constituency parser is selected as an approach due to its pros of easy to understand.
- To remove ambiguity PCFG is used as it provides partial solution to generate ambiguity free tree. Unlexicalized PCFG is used as it provides more accuracy and no annotation overhead.
- Result is based on genia tagger and Penn tree bank POS tags. Penn tree bank is basically used for English Domain. Genia tagger is helpful for biomedical text.
- Approximity of tree is calculated through which probability of accuracy of generating tree is defined.
- Constituency Parser can be improvised for dependency parser
- Constituency clinical parser can be generated in other languages such as Hindi, Chinese, Spanish etc.
- Lexicalised parsing can also implement.

References

[1] C. Friedman, "Towards a comprehensive medical language processing system: Methods and issues," *Proceedings of the AMIA Annual Fall Symposium: American Medical Informatics Association*, pp. 595- 599, 1997.

[2] Koehler, and B. Spencer, "SymText: A natural language understanding system for encoding free text medical data," Doctor Dissertation, University of Utah, 1998.

[3] L. M. Christensen, P. J. Haug, and M. Fiszman, "MPLUS: A probabilistic medical language understanding system," *Proceedings of the Workshop on Natural Language Processing in the Biomedical Domain: Association for Computational Linguistics*, pp. 29-36, 2002.

[4] B. Hazlehurst, H. R. Frost, D. F. Sittig, and V. J. Stevens, "MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record," *Journal of the American Medical Informatics Association*, vol. 12, no. 5, pp. 517-529, 2005.

[5] S. Goryachev, M. Sordo, and Q. T. Zeng, "A suite of natural language processing tools developed for the I2B2 project," *AMIA Annual Symposium Proceedings*, p. 931, 2006.

[6]G. K. Savova, J. J. Masanz, P. V. Ogren, J. Zheng, S. Sohn, K. C. Kipper-Schuler, and C. G. Chute, "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications," *Journal of the American Medical Informatics Association*, vol. 17, no. 5, pp. 507-513, 2010.

[7] A. R. Aronson, and F.-M. Lang, "An overview of MetaMap: Historical perspective and recent advances," *Journal of the American Medical Informatics Association*, vol. 17, no. 3, pp. 229-236, 2010.

[8] H. Xu, S. P. Stenner, S. Doan, K. B. Johnson, L.R. Waitman, and J. C. Denny, "MedEx: A medica- tion information extraction system for clinical narratives," *Journal of the American Medical Informatics Association*, vol. 17, no. 1, pp. 19-24, 2010.

[9]V. Garla, V. L. Re, Z. Dorey-Stein, F. Kidwai, M. Scotch, J. Womack, A. Justice, and C. Brandt, "The Yale cTAKES extensions for document classification: Architecture and application," *Journal of the American Medical Informatics Association*, vol. 18, no. 5, pp. 614- 620, 2011.

[10] Y. K. Lin, H. Chen, and R. A. Brown, "MedTime: A temporal information extraction system for clinical narratives," *Journal of Biomedical Informatics*, pp. 20-28, 2013.

[11] S. Sohn, C. Clark, S. R. Halgrim, S. P. Murphy, C. G. Chute, and H. Liu, "MedXN: An open source medication extraction and normalization tool for clinical text," *Journal of the American Medical Informatics Association*, vol. 21, no. 5, pp. 858-865, 2014.

[12] Foster R. Goss, Joseph M. Plasek, Jason J. Lau, Diane L. Seger, Frank Y. Chang, and Li Zhou, An Evaluation of a Natural Language Processing Tool for Identifying and Encoding Allergy Information in Emergency Department Clinical Notes, AMIA Annu Symp Proc.2014, pp. 580–588.

[13] Hua Xu, Samir Abdel Rahman, Min Jiang, Jung-wei Fan and Yang Huang, An Initial Study of Full Parsing of Clinical Text using the Stanford Parser. In IEEE International Conference on Bioinformatics and Biomedicine Workshops,2011, pp. 607-614.

[14] Finkel, J., S. Dingare, and H. Nguyen, Exploiting context for biomedical entity recognition: From syntax to the web. In Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its applications, 2004.

[15] Fundel K., R. Kuffer, and R. Zimmer, RelEx-Relation extraction using dependency parse tree. Bioinformatics, 2007. 23(3): pp. 365-371.

[16] Huang Y. et al., Improved identification of noun phrases in clinical radiology reports using a high-performance statistical natural language parser augmented with the UMLS specialist lexicon. J Am Med Inform Assoc., 2005. 12(3): pp. 275-85.

[17] Bikel, D.M., On the parameter space of generative lexicalized statistical parsing models. 2004.

[18] Charniak, E. and M. Johnson, Coarse-to-fine n-best parsing and Maxent discriminative reranking, in Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics: Ann Arbor., Michigan,2005, pp. 173-180.

[19] McClosky, D., E. Charniak, and M. Johnson, Effective self-training for parsing, in Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. Association for Computational Linguistics: New York,2006, pp. 152-159.

[20] Rohit J Kate, Unsupervised grammar induction of clinical report sublanguage Biomed semantics 2012.3(Suppl 3): S4

[21] Gabriel J. Ferrer, An Interactive Parser Generator for context-free grammars. In ACM Digital Library Journal of Computing Sciences in Colleges, Volume 28 Issue 2,2012, pp. 103-110

[22] Alessandro Warth and Tony garnock, Modular Semantic Actions. In ACM digital Library ,2016, pp. 108-119.

[23] Ricky K. Taira and Stephen G. Soderland, A Statistical Natural Language Processor for Medical Report, AMIA Proc.1999, pp. 970-974.

[24] Friedrich Steinmann, Dependency Parsing for Medical Language and Concept Representation, Artif Intell Med.,1998 Jan;12(1):77-86.

[25] Ernestina Menaselvas et al., Clinical Narrative Analytical Challenges, Springer International Publishing,2016, pp. 23-32

[26] P.O. El Guedj and P. Nugues, A Chart Parser to analyse large medical corpora, IEEE Xplore Digital Library,2002.

[27] Nuala A. Bennett, Qin He, Kevin Powell, Bruce R. Schatz, Extracting Noun Phrases for all of MEDLINE,1999, AMIA Inc., pp. 671-675.

[28] MingJiang, Jana Diesner, A Constituency Parsing Tree based method for Relation Extraction from Abstracts of Scholarly Publications, Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13),2019, pp. 186–191

Bibliography

- http://nactem7.mib.man.ac.uk/geniatagger/index.html
- http://nlp.stanford.edu/software/lex-parser.shtml
- https://www.nlm.nih.gov/research/umls/index.html
- https://www.nlm.nih.gov/healthit/snomedct/index.html